

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 March 2003 (27.03.2003)

PCT

(10) International Publication Number  
**WO 03/025148 A2**

(51) International Patent Classification<sup>7</sup>: C12N

(21) International Application Number: PCT/US02/29964

(22) International Filing Date:  
19 September 2002 (19.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/323,739 19 September 2001 (19.09.2001) US  
Not furnished 13 September 2002 (13.09.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US Not furnished (CIP)  
Filed on Not furnished

(71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). ASUNDI, Vinod [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). GOODRICH, Ryle, W. [US/US]; 3461 Saint Susan Place, Los Angeles, CA 90066 (US). REN, Feiyan [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). ZHANG, Jie [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). ZHAO, Qing, A. [CN/US]; 1556 Kooser Rd, San Jose, CA 95118 (US). WANG, Jian-Rui [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). GHOSH, Malabika [IN/US]; 620 Iris Avenue, Apt. 133, Sunnyvale, CA 94086 (US). XUE, Aidong, J. [US/US]; 1621 S. Mary Avenue, Sunnyvale, CA 94087 (US). WEHRMAN, Tom

[US/US]; CCSR Mol Pharm 3210, 269 W. Campus Dr, Stanford, CA 94305 (US). WENG, Gezhi [US/US]; 415 Moraga Avenue, Piedmont, CA 94611 (US). ZHOU, Ping [US/US]; 7595 Newcastle Drive, Cupertino, CA 95014 (US). DRMANAC, Radoje, T. [US/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US). WANG, Dunrui [CN/US]; 12252 Pepper Tree Lane, Poway, CA 92064 (US). HALEY-VICENTE, Dana [US/US]; 1305 Eagle Glen, Escondido, CA 92029 (US).

(74) Agents: RIN-LAURES, Li-Hsien et al.; Hyseq, Inc., 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.



WO 03/025148 A2

## NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

### 1. CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Application Serial No. 60/323,739 filed September 19, 2001 entitled "Novel Nucleic Acids and Polypeptides", Attorney Docket No. 809, which is a continuation-in-part application of PCT Application Serial No. PCT/US00/35017 filed December 22, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP3A/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/552,317 filed April 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/488,725 filed January 21, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784; PCT Application Serial No. PCT/US01/02623 filed January 25, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/491,404 filed January 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785; PCT Application Serial No. PCT/US01/03800 filed February 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/560,875 filed April 27, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/496,914 filed February 03, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787; PCT Application Serial No. PCT/US01/04927 filed February 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,409 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/515,126 filed February 28, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788; PCT Application Serial No. PCT/US01/04941 filed March 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/574,454 filed May 19, 2000 entitled "Novel Contigs Obtained from Various

Libraries", Attorney Docket No. 789CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/519,705 filed March 07, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789; PCT Application Serial No. PCT/US01/08631 filed March 30, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/649,167 filed August 23, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/540,217 filed March 31, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790; PCT Application Serial No. PCT/US01/08656 filed April 18, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/770,160 filed January 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP, which is in turn a continuation-in-part application of U.S. Application Serial No. 09/552,929 filed April 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791; and PCT Application Serial No. PCT/US01/14827 filed May 16, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,408 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792; all of which are incorporated herein by reference in their entirety.

## **2. BACKGROUND OF THE INVENTION**

### **2.1 TECHNICAL FIELD**

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

### **2.2 BACKGROUND**

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, circulating soluble factors, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression

cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence  
5 cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader  
10 sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types  
15 of data and products dependent on DNA and amino acid sequences.

### 3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules,  
20 cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including  
25 expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by  
30 hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-276, or 553-772 and are provided in



the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases or unknown. In the amino acids provided in the Sequence Listing, \* corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences  
5 that hybridize to the complement of SEQ ID NO: 1-276, or 553-772 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-276, or 553-772. A polynucleotide comprising a nucleotide sequence having at least  
10 90% identity to an identifying sequence of SEQ ID NO: 1-276, or 553-772 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-276, or 553-772. The sequence information can be a segment of any one of SEQ ID NO: 1-276, or 553-772 that uniquely  
15 identifies or represents the sequence information of SEQ ID NO: 1-276, or 553-772.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The  
20 array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences  
25 (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their  
30 chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-276, or 553-772 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the

nucleic acid sequences of SEQ ID NO: 1-276, or 553-772 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., *Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

5           The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-276, or 553-772; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-276, or 553-772; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-276, or 553-772. The  
10   polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-276, or 553-772; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in SEQ ID NO: 1-276, or 553-772; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a  
15   polynucleotide which encodes a species homologue (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in SEQ ID NO: 277-552, or 773-992, or Tables 3, 4A, 4B, 5, 6, or 8.

          The isolated polypeptides of the invention include, but are not limited to, a polypeptide  
20   comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-276, or 553-772; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization  
25   conditions. Biologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically  
30   engineered cells (e.g. host cells) of the invention.

          The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such processes is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g.*, *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions.

- 5 The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a
- 10 method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

- The invention also provides kits comprising polynucleotide probes and/or
- 15 monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

- The invention also provides methods for the identification of compounds that
- 20 modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The
- 25 invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected
- 30 the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for

treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can affect such modulation either on the level of target gene/protein expression or target protein activity.

5           The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Tables 2A and 2B); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Tables 4A and 4B). If no homology is set forth for a sequence,  
10       then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

#### 4. DETAILED DESCRIPTION OF THE INVENTION

##### 15           4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

          The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the  
20       invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

25           The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

          The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the  
30       complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only certain portion(s) of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded

molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G, or T (U) or unknown. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil).

Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is

capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of  
5 nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most  
10 preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related  
15 parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NO: 1-276, or 553-772.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal  
20 DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in  
25 Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-276, or 553-772. The sequence information can be a segment of any one of SEQ ID NO: 1-276, or 553-772 that  
30 uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-276, or 553-772, or those segments identified in Tables 3, 4A, 4B, 5, 6, or 8. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three

billion base pairs in one set of chromosomes. Because  $4^{20}$  possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes.

Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for

5 expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human  
10 genome with a single mismatch is calculated by multiplying the probability for a full match ( $1/4^{25}$ ) times the increased probability for mismatch at each nucleotide position ( $3 \times 25$ ). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

15 The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence.

20 While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is  
25 restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7  
30 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids.



Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full-length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e.g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be

reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (e.g., microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or

elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

5           The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that  
10   are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (*e.g.* Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (*e.g.* Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

15           Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

          The term "stringent" is used to refer to conditions that are commonly understood in  
20   the art as stringent. Stringent conditions can include highly stringent conditions (*i.e.*, hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (*i.e.*, washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

25           In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

          As used herein, "substantially equivalent" or "substantially similar" can refer both to  
30   nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of

those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have

5 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than

10 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least

15 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at

20 least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of

25 determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a new stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun Hein method (Hein, J. (1990) *Methods Enzymol.* 183:626-645). Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

30 The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal

integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

5 As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid  
10 molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

#### 15 4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-276, or 553-772; a polynucleotide encoding any  
20 one of the peptide sequences of SEQ ID NO: 1-276, or 553-772; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-276, or 553-772. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID  
25 NO: 1-276, or 553-772; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing, or Table 8; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homologue of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO:  
30 277-552, or 773-992 (for example, as set forth in Tables 3, 4A, 4B, 5, 6, or 8). Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable

immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

5 The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

10 The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 15 1-276, or 553-772 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-276, or 553-772 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-276, or 553-772 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

20 The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpr, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

25 The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, e.g., at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, 30 and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide

sequences of SEQ ID NO: 1-276, or 553-772, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-276, or 553-772, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-276, or 553-772 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology results for the nucleic acids of the present invention, including SEQ ID NO: 1-276, or 553-772 can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST (Basic Local Alignment Search Tool) program is used to search for local sequence alignments (Altschul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using FASTXY algorithm may be performed.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be



prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA

fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., *supra*, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention could be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-276, or 553-772, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient

restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

5       The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-276, or 553-772 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-276, or 553-  
10   772 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention.  
15   The following vectors are provided by way of example: Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

      The isolated polynucleotide of the invention may be operably linked to an expression  
20   control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means  
25   that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

      Promoter regions can be selected from any desired gene using CAT  
30   (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse

metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK),  $\alpha$ -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., Nat. Biotech 17, 870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

### 4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-276, or 553-772, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1-276, or 553-772 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-276, or 553-772 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-276, or 553-772, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more

preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)*w*, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the

case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target  
5 selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve  
10 sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific  
15 double-stranded hybrids with complementary RNA in which, contrary to the usual  $\alpha$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

#### 4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a  
25 complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1-276, or 553-772). For example, a derivative of  
30 *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease

activity from a pool of RNA molecules. See, e.g., Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (e.g., promoter and/or enhancers) to form triple  
5 helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability,  
10 hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural  
15 nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

20 PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes  
25 when used in combination with other enzymes, e.g., S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by  
30 the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA



portion while the PNA portion would provide high binding affinity and specificity.

PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996)

above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup

5 (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to  
10 produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such

15 as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents  
20 (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

#### 25 4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically  
30 engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory*

Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene

5 targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, and regulatory protein binding sites or combinations of said sequences.

10 Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA

15 molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory

20 element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are

25 contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting

30 sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. 5 PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

#### 4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a  
10 polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 277-552, or 773-992 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-276, or 553-772 or the corresponding full length or mature protein.  
Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the  
15 nucleotide sequences set forth in SEQ ID NO: 1-276, or 553-772 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 277-552, or 773-992 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as  
20 SEQ ID NO: 277-552, or 773-992 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological  
25 activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 277-552, or 773-992.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as  
30 described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as

immunoglobulins for many purposes, including increasing the valency of protein binding sites. Fragments are also identified in Tables 3, 4A, 4B, 5, 6, or 8.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein  
5 coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The predicted signal sequence is set forth in Table 6. The mature form of such protein may be obtained and confirmed by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell and sequencing of the cleaved product. One of skill in the art will recognize that the actual cleavage site may be different  
10 than that predicted in Table 6. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed (See, e.g.,  
15 Sakal et al., Prep. Biochem. Biotechnol. (2000), 30(2), pp. 107-23, incorporated herein by reference).

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic  
20 acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the  
25 ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or  
30 tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may

be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified  
5 from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic  
10 sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the  
15 methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments  
20 include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated  
25 polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include  
30 fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then  
5 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that  
10 are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 277-552, or 773-992.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the  
15 protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of  
20 interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S.  
25 Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for  
30 biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.



Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

5       The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described  
10       in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

      The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting  
15       expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™;  
20       one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

      Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as  
25       a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

30       Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide

a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces  
5 fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability.  
10 Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic  
15 agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

#### 4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE

##### 20 IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group,  
25 University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by  
30 reference), Pfam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), the GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95,

13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) incorporated herein by reference).

- 5 Polypeptide sequences were examined by a proprietary algorithm, SeqLoc that separates the proteins into three sets of locales: intracellular, membrane, or secreted. This prediction is based upon three characteristics of each polypeptide, including percentage of cysteine residues, Kyte-Doolittle scores for the first 20 amino acids of each protein, and Kyte-Doolittle scores to calculate the longest hydrophobic stretch of the said protein. Values of
- 10 predicted proteins are compared against the values from a set of 592 proteins of known cellular localization from the Swissprot database (<http://www.expasy.ch/sprot>). Predictions are based upon the maximum likelihood estimation.

Presence of transmembrane region(s) was detected using the TMPred program ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)).

- 15 The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

#### 4.7 CHIMERIC AND FUSION PROTEINS

- 20 The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the
- 25 invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

- 30 For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

5 In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*.  
10 The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a  
15 subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional  
20 techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of  
25 gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety  
30 (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

#### 4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be

modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences.

- 5 See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with
- 10 the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

- In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control
- 15 of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment
- 20 regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion
- 25 properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

- The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple
- 30 deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are

deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

#### 4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by  
5 supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to  
10 express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or  
15 inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals,  
20 can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No  
25 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or  
30 even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.



#### 4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention

5 may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment.

10 Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target

15 gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular

20 antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

##### 25 4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage

30 of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA

sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

#### 4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid

preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

#### 5           4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations.

- 10 A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell
- 15 proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

- Assays for T-cell or thymocyte proliferation include without limitation those
- 20 described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology
- 25 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

- Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan
- 30 eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- $\gamma$ , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

#### 4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotent or pluripotent state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors.

The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies

would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

*In vitro* cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. *Proc. Natl. Acad. Sci, U.S.A.*, 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the

invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

#### 4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

5 A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, 10 thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in 15 supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and 20 therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or 25 heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

30 Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

- Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994;
- 5 Hirayama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*.
- 10 R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc.,
- 15 New York, N.Y. 1994.

#### 4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and

20 tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have

25 prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming

30 cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast



activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from

chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

#### **4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY**

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and

disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also be useful in the treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., *Toxicology* 125: 59-66, 1998), skin prick test (Hoffmann et al., *Allergy* 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., *Arch. Toxicol.* 73: 501-9), and murine local lymph node assay (Kimber et al., *J. Toxicol. Environ. Health* 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of

an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing  
5 non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without  
10 limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition  
15 as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may  
20 avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in  
25 humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed.,  
30 Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and  $\beta_2$  microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro

antibody production, Mond, J. J. and Brunswick, M. In *Current Protocols in Immunology*. J. E. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnolli et al., *J. Immunol.* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., *J. Immunol.* 134:536-544, 1995; Inaba et al., *Journal of Experimental Medicine* 173:549-559, 1991; Macatonia et al., *Journal of Immunology* 154:5071-5079, 1995; Porgador et al., *Journal of Experimental Medicine* 182:255-260, 1995; Nair et al., *Journal of Virology* 67:4062-4069, 1993; Huang et al., *Science* 264:961-965, 1994; Macatonia et al., *Journal of Experimental Medicine* 169:1255-1264, 1989; Bhardwaj et al., *Journal of Clinical Investigation* 94:797-807, 1994; and Inaba et al., *Journal of Experimental Medicine* 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., *Cytometry* 13:795-808, 1992; Gorczyca et al., *Leukemia* 7:659-670, 1993; Gorczyca et al., *Cancer Research* 53:1945-1951, 1993; Itoh et al., *Cell* 66:233-243, 1991; Zacharchuk, *Journal of Immunology* 145:4037-4045, 1990; Zama et al., *Cytometry* 14:891-897, 1993; Gorczyca et al., *International Journal of Oncology* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., *Blood* 84:111-117, 1994; Fine et al., *Cellular Immunology* 155:111-122, 1994; Galy et al., *Blood* 85:2770-2778, 1995; Toki et al., *Proc. Nat. Acad. Sci. USA* 88:7548-7551, 1991.

#### 4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in:

Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

#### **4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY**

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to



tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margules, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

#### 4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis

Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

#### 4.10.11 CANCER DIAGNOSIS AND THERAPY

5 Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing  
10 malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation,  
15 inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies  
20 including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal  
25 neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central  
30 nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Kaposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention

(including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

*In vitro* models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987)

Culture of Animal Cells: A Manual of Basic Technique, Wiley-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

#### 4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

5       Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182  
10       (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14. Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

#### 15       4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One  
20       method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or  
25       fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries  
30       comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product  
5 libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide  
10 and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol.* 9(3):205-23 (1998); Hruby  
15 et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dörner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then  
20 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The  
25 toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

#### 4.10.14 ASSAY FOR RECEPTOR ACTIVITY

30 The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening

assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of  
5 compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does  
10 not. The responses of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic  
15 chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the  
20 extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify  
25 signaling molecules involved in receptor activity.

#### 4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in  
30 the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an

inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis,

- 5 complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis,
- 10 acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to
- 15 intrauterine infections.

#### **4.10.16 LEUKEMIAS**

- Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of
- 20 the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

25

#### **4.10.17 NERVOUS SYSTEM DISORDERS**

- Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of
- 30 therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include



but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

- (i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or  
5 compression injuries;
- (ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;
- (iii) infectious lesions, in which a portion of the nervous system is destroyed or  
10 injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;
- (iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration  
15 associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;
- (v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency,  
20 Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;
- (vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;
- (vii) lesions caused by toxic substances including alcohol, lead, or particular  
25 neurotoxins; and
- (viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various  
30 etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival

or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- 5 (iii) increased production of a neuron-associated molecule in culture or *in vivo*,  
*e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method  
10 set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons  
may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or  
Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of  
neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody  
binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor  
15 neuron dysfunction may be measured by assessing the physical manifestation of motor  
neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor  
20 neurons as well as other components of the nervous system, as well as disorders that  
selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited  
to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis,  
infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-  
Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory  
25 Neuropathy (Charcot-Marie-Tooth Disease).

#### 4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing,  
30 infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites;  
effecting (suppressing or enhancing) bodily characteristics, including, without limitation,  
height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or  
organ or body part size or shape (such as, for example, breast augmentation or diminution,

change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s);  
5 effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of  
10 the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

#### 4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential  
20 predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in  
25 humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate  
30 fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that

hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

#### 4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et al., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

#### 4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

5

#### 4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01  $\mu\text{g/kg}$  to 100  $\text{mg/kg}$  of body weight, with the preferred dose being about 0.1  $\mu\text{g/kg}$  to 10  $\text{mg/kg}$  of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

#### 4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other

materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration.

The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF- $\alpha$  and TGF- $\beta$ ), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound

sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

#### 4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in

fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

#### 4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical



composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired,

disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such

as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the  
5 preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or  
10 other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion  
15 exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate  
20 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics.  
25 Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds  
30 may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable

matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides,

diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like.

Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

- 5           The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient.
- 10          Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to
- 15          practice the method of the present invention should contain about 0.01  $\mu$ g to about 100 mg (preferably about 0.1  $\mu$ g to about 10 mg, more preferably about 0.1  $\mu$ g to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition
- 20          topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically
- 25          useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active
- 30          ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, 5 tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised 10 of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having 15 diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, 20 ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful 25 herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, 30 proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet

derived growth factor (PDGF), transforming growth factors (TGF- $\alpha$  and TGF- $\beta$ ), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

#### 4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be

estimated initially from appropriate *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the  $IC_{50}$  as  
5 determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic  
10 efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the  $LD_{50}$  (the dose lethal to 50% of the population) and the  $ED_{50}$  (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between  $LD_{50}$  and  $ED_{50}$ . Compounds which exhibit high therapeutic  
15 indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the  $ED_{50}$  with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of  
20 administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in*  
25 *vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of  
30 the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.



An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01  $\mu\text{g/kg}$  to 100  $\text{mg/kg}$  of body weight daily, with the preferred dose being about 0.1  $\mu\text{g/kg}$  to 25  $\text{mg/kg}$  of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at  
5 longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

#### 10 4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a  
15 compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

#### 20 4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,  
25  $\text{F}_{\text{ab}}$ ,  $\text{F}_{\text{ab}}$ , and  $\text{F}_{(\text{ab})_2}$  fragments, and an  $\text{F}_{\text{ab}}$  expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a  
30 reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in

5 SEQ ID NO: 277-552, or 773-992, or Tables 3, 4A, 4B, 5, 6, or 8, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope.

Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues.

10 Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of

15 a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and

20 Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog

25 thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence

30 identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine

binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), *Antibodies A Laboratory Manual*; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known

in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

#### 4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8  
5 (April 17, 2000), pp. 25-28).

#### 4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular  
10 species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

15 Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256, 495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be  
20 immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell  
25 line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that  
30 preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107, 220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as

a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA  
5 also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted  
10 for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

#### 4.13.3 HUMANIZED ANTIBODIES

15 The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab',  
20 F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321, 522-525 (1986); Riechmann et al., Nature, 332, 323-327 (1988); Verhoeven et al., Science, 239, 1534-1536 (1988)), by substituting  
25 rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise  
30 substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion

of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2, 593-596 (1992)).

#### 4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80, 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227, 381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368, 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al, (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13, 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains



in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then  
5 obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after  
10 immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for  
15 example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent  
20 rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

25 A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The  
30 hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

#### 4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

5 According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of  $F_{ab}$  expression libraries (see e.g., Huse, et al., 1989 Science 246, 1275-1281) to allow rapid and effective identification of monoclonal  $F_{ab}$  fragments with the desired specificity for a protein or  
10 derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an  $F_{(ab)2}$  fragment produced by pepsin digestion of an antibody molecule; (ii) an  $F_{ab}$  fragment generated by reducing the disulfide bridges of an  $F_{(ab)2}$  fragment; (iii) an  $F_{ab}$  fragment generated by the treatment of the antibody molecule with papain and a reducing  
15 agent and (iv)  $F_v$  fragments.

#### 4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of  
20 the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two  
25 immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305, 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished  
30 by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10, 3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121, 210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g.  $F(ab')_2$  bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229, 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate  $F(ab')_2$  fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The  $Fab'$  fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the  $Fab'$ -TNB derivatives is then reconverted to the  $Fab'$ -thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other  $Fab'$ -TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally,  $Fab'$  fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175, 217-225 (1992) describe the production of a fully humanized bispecific antibody  $F(ab')_2$  molecule. Each  $Fab'$  fragment was separately secreted from *E. coli* and subjected to directed chemical

coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5), 1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90, 6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the  $V_H$  and  $V_L$  domains of one fragment are forced to pair with the complementary  $V_L$  and  $V_H$  domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152, 5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147, 60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG ( $Fc\gamma R$ ), such as  $Fc\gamma RI$  (CD64),  $Fc\gamma RII$  (CD32) and  $Fc\gamma RIII$  (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen.

Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

#### 4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

#### 4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176, 1191-1195 (1992) and Shopes, J. Immunol., 148, 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53, 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3, 219-230 (1989).

#### 4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

#### 4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the

presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-276, or 553-772 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-276, or 553-772 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the

present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means  
5 having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present  
10 invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target  
15 sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available  
20 algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The  
25 most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally  
30 selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,



but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

#### 4.15 TRIPLE HELIX FORMATION

5 In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple  
10 helix-see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 15241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization  
15 blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

#### 20 4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

25 In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization  
30 conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

5 In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods  
10 employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science  
15 Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or  
20 membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is  
25 compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies  
30 of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

#### 4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-276, or 553-772, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and  
5 detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting  
10 the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression  
15 of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to  
20 activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in  
25 the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

30 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"

when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al.,

- 5 Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

- In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or  
10 EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple  
15 helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

- Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix -  
20 see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Okano, J. Neurochem. 56, 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks  
25 translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

- Agents which bind to a protein encoded by one of the ORFs of the present invention  
30 can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

#### 4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of  
5 the nucleotide sequences SEQ ID NO: 1-276, or 553-772. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NO: 1-276, or 553-772 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ  
10 hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related  
15 genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or  
20 SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well-known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage  
25 analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

30 Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal

map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

#### 5           4.20    **PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES**

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6), 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8), 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridgeheads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen *et al.*, (1991). In this technology, a phosphoramidate bond is employed (Chu *et al.*, (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins

the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ $\mu$ l) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-Melm<sub>7</sub>), is then added to a final concentration of 10 mM 1-Melm<sub>7</sub>. A ss DNA solution is then dispensed into CovaLink NH strips (75  $\mu$ l/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-Melm<sub>7</sub>, is made fresh and 25  $\mu$ l added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) *Science* 251(4995), 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) *Nucleic Acids Res.*, 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) *Anal. Biochem.* 169(1), 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.



One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Nat'l. Acad. Sci., USA 91(11), 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

#### 4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schrieffer *et al.* (1990) Nucleic Acids Res. 18(24), 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, CviJI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *CviJI* normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*CviJI*\*\*), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *CviJI*\*\* digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *CviJI*\*\* restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 µg instead of 2-5 µg); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

#### 4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm<sup>2</sup>, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be

spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm<sup>2</sup> and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

## 5.0 EXAMPLES

### 5.1 EXAMPLE 1

#### Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences.

## 5.2 EXAMPLE 2

### Assemblage of Novel Nucleic Acids

The contigs or nucleic acids of the present invention, designated as SEQ ID NO: 553-772 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene, and exons from public domain genomic sequences predicated by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above databases that would extend the assemblage. Further, inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

The novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (SEQ ID NO: 553-772) of the present invention, and their corresponding translation start and stop nucleotide locations to each of SEQ ID NO: 553-772 were obtained using one of two methods. Polypeptides were obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of the translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 (1990), herein incorporated by reference). Alternatively, polypeptides were obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268:78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame.

## 5.3 EXAMPLE 3

### Novel Nucleic Acids

The novel nucleic acids of the present invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequences were checked using FASTY and/or BLAST against Genbank (i.e., dbEST, gb pri, UniGene, and Genpept) and the Geneseq (Derwent). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NO: 1-552.

The nucleic acid sequences of the present invention were confirmed to have at least one transmembrane domain using the TMPred program ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html), herein incorporated by reference).

Table 1 shows the various tissue sources of SEQ ID NO: 1-276.

The homologs for polypeptides SEQ ID NO: 277-552, that correspond to nucleotide sequences SEQ ID NO: 1-276 were obtained by a BLASTP search against Genpept release 124 and Geneseq (Derwent) release 200117 and against Genpept release 129 and Geneseq (Derwent) release (July 18, 2002). The results showing homologues for SEQ ID NO: 277-552 from Genpept 124 are shown in Table 2A. The results showing homologues for SEQ ID NO: 277-552 from Genpept 129 are shown in Table 2B.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6, 219-235 (1999), <http://motif.stanford.edu/ematrix-search/> herein incorporated by reference), all the polypeptide sequences were examined to determine

whether they had identifiable signature regions. Scoring matrices of the eMatrix software package are derived from the BLOCKS, PRINTS, PFAM, PRODOM, and DOMO databases. Table 3 shows the accession number of the homologous eMatrix signature found in the indicated polypeptide sequence, its description, and the results obtained which include  
5 accession number subtype; raw score; p-value; and the position of signature in amino acid sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4A shows the  
10 name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence as described in United States priority application serial number 60/323,739, filed September 19, 2001, herein incorporated by reference in its entirety. Table 4B shows the name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence using Pfam version 7.2.

15 Further description of the Pfam models can be found at <http://pfam.wustl.edu/>.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ ID NO: 1-276 (i.e. SEQ ID NO: 277-552). Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by  
20 Altschul et al, (Nucl. Acids. Res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA,) which is an automated sequence and structure searching procedure (<http://www.msi.com/>), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with  
25 the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows: "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as  
30 annotated by the PDB files (<http://www.rcsb.org/PDB/>); start and end amino acid position of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score is produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David

Eisenberg's laboratory (US patent no. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:13597-12502. The verify score produced by GeneAtlas normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

Verify score (normalized) = (raw score – 1/2 high score)/(1/2 high score)

The PFM score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potentials (MFP). As given in Table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good model. Similarly, a PMF score between 0 to 1.0, with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may also be determined by one of skill in the art based all the information in Table 5 taken in totality.

Table 6 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences.

Table 7 correlates each of SEQ ID NO: 1-276 to a specific chromosomal location.

Table 8 shows the number of transmembrane regions, their location(s), and TMPred score obtained, for each of the SEQ ID NO: 277-552 that had a TMPred score of 500 or greater, using the TMPred program ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)).

Table 9 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-276, their corresponding polypeptide sequences SEQ ID NO: 277-552, their corresponding

priority contig nucleotide sequences SEQ ID NO: 553-772, their corresponding priority contig polypeptide sequences SEQ ID NO: 773-992, and the US serial number of the priority application (all of which are herein incorporated in their entirety), in which the contig sequence was filed.

- 5        Table 10 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-276, the novel polypeptide sequences SEQ ID NO: 277-552, and the corresponding SEQ ID NO in which the sequence was filed in priority US application bearing serial number 60/323,739, filed September 19, 2001.



116  
Table 1

Tissue origin	Library/RNA source	HYSEQ Library Name	SEQ ID NO:
adult brain	GIBCO	AB3001	8 76 78 80 101-102 109-111 113 153 194 205 265
adult brain	GIBCO	ABD003	1-3 8-9 11 14 23 29 41 76 78 84 89 93 95 104-106 109-111 113-114 126-127 136-139 151-152 162 164-166 176 178 181 211 224 263
adult brain	Clontech	ABR001	23 38-39 47 91 103 106 139 143 171 224 235 244
adult brain	Clontech	ABR006	1-3 8-9 22 29-30 36 38-39 41 51-53 66 76 79 88 91 93 101-102 113 121 123 126-127 133-134 139 147 161-162 170 186 192 198 202-203 211 219 221 225 232 234 252 262-263 271 275
adult brain	Clontech	ABR008	1-3 6 9-11 13 15 24 30-31 33 36 38-39 41 44 46-47 55-56 61-65 74 76 80-81 87 93 95 99-102 104-106 109-110 114-115 122-123 127-128 138-140 143 154-155 164-167 169-170 172-174 178 186 188 190 199-200 202-206 211 213 217-219 221-222 230 232 234 242-243 245 252 263 271 276
adult brain	BioChain	ABR012	5 28 161 211
adult brain	BioChain	ABR013	144 154
adult brain	Invitrogen	ABR014	76 115
adult brain	Invitrogen	ABR015	13 15 178 211
adult brain	Invitrogen	ABR016	37 95 101-102
adult brain	Invitrogen	ABT004	6 23 47 79 101-103 106 109-110 113 115 137 154 158 171-173 176 189-190 192-193 199 231 269 271
cultured preadipocytes	Stratagene	ADP001	4 26 33 81-83 86 99-102 114-115 132 154 181 193
adrenal gland	Clontech	ADR002	9 13 32 40-41 57 72 76 84 93 103-105 115 120 122 126 133 138 140 155 157 164-166 171 187 194 199-200 209 211 220 224-225 264
adult heart	GIBCO	AHR001	1-3 5-6 8 11-12 14 21 26 28 41 55 87 99-104 106 109-110 113 115 118 120 124-125 132 136 139 145 153-154 158 160 169 180 195 198 200 211 253 267
adult kidney	GIBCO	AKD001	1-7 15-16 19-21 28 42 57 60 84 87 91 95 101-102 104-105 107 113 115 121-123 126 129 132-133 137-138 140-144 149 151-152 155-156 159 163-167 178 194 198 205 211 213 230 235 242 253 261 265
adult kidney	Invitrogen	AKT002	1-4 6 15 20-21 41 43 45-46 60 90 101-102 105-106 108 111 114-115 121 134 137 143 151-154 157 163 178 198 205 213 223-224 230 246 265
adult lung	GIBCO	ALG001	5 24 72 78 136 158 164-166 168 267 270
lymph node	Clontech	ALN001	64 121 154 216 235
young liver	GIBCO	ALV001	1-3 5 28 101-102 104 122 125 132 164-166 172 178 201 213 220 224
adult liver	Invitrogen	ALV002	15-16 26 42 47 51-53 58 60 75 84 87 101-102 104 109-110 112 114-115 138 143 154 164-166 172 178 195 199 207 236

117  
Table 1

Tissue origin	Library/RNA source	HYSEQ Library Name	SEQ ID NO:
			252 254
adult liver	Clontech	ALV003	1-3 104 115 120 169 172
adult ovary	Invitrogen	AOV001	1-5 21-22 26 28-29 32 38-39 41 48 78 84 86-87 95 99-102 104 106-111 113-115 118 120-121 126 131-134 136 138 145- 146 149-150 153-154 157-158 160 163 168-171 180 186-188 192 194 198-199 201 209 211 214 216 224-225 231 242 246 253 265
adult placenta	Clontech	APL001	16 46 136
placenta	Invitrogen	APL002	4 26 47 60 101-102 109-110 143 153 164- 166 178 242
adult spleen	GIBCO	ASP001	1-3 6 15 17 72 82-83 101-102 104 109- 110 118 121 129 132 136 158 178 181 198 238 240
adult testis	GIBCO	ATS001	1-3 6 13 21 60 80 137 145 150 158 171 247
adult bladder	Invitrogen	BLD001	6 94 114 164-166 169 178 188 190 200 252
bone marrow	Clontech	BMD001	1-3 11-14 29 86 99-100 103-106 111 113 121-124 134 147-148 197-198 211 213 225 230 253-254 264
bone marrow	GF	BMD002	6 9 13 22 32 51-53 55 60 74 82-83 93 95 99-105 108-110 113 122-123 129 131 139 143 147 153 159 161 164-166 178 186 190 211 221 224 230 234 246 248 250 253-254
adult colon	Invitrogen	CLN001	47 60 158 173 181 201 211
adult cervix	BioChain	CVX001	1-3 8 14 29 38-39 41-42 51-53 72 78-80 84 86-87 97 99-100 104 106-107 111 113 115 121-122 124 132-134 136 138 143 145 153-155 178 181 188 195 198-199 209 211 223 225 240 242 252-253 267
diaphragm	BioChain	DIA002	182
endothelial cells	Stratagene	EDT001	4-5 15-16 26 28-29 36 47 51-53 57 60 78 99-102 104-105 107 109-110 113 115 121 123 131-132 136 138 144 150 154 158 164-166 171 178 198 201 213 224 235 251-252
fetal brain	Clontech	FBR001	1-3 31 42 76 79 137 154
fetal brain	Clontech	FBR004	36 79 154
fetal brain	Clontech	FBR006	5 10-11 13 15 24-25 30-33 38-39 41-42 47 62-64 76 78 80-81 95 99-102 104-105 109-110 115 117-118 122-123 126-128 131 133 138 143 147 154 167 173 175 178 188 194 199-200 202-204 206-207 211 218 222 234-235 244-245 252 262 266 271-272 275
fetal brain	Clontech	FBRs03	5 28
fetal brain	Invitrogen	FBT002	6 15 24 35-36 41 64 101-102 113 127 137 144 153-154 162 178 192 194 216
fetal heart	Invitrogen	FHR001	6 14-15 21 30 46 51-53 68 80-81 87 95 101-102 106-107 109-110 113 115 118 122 136 139 145 178 188 196-197 199- 201 211 214 253 256-257 261

118  
Table 1

Tissue origin	Library/RNA source	HYSEQ Library Name	SEQ ID NO:
fetal kidney	Clontech	FKD001	1-3 6 105 109-110 178 198 265
fetal kidney	Clontech	FKD002	10 46 57 107 113 118 154-155 161 186 205 221 253 267
fetal lung	Clontech	FLG001	9 13 121 132 136 161 181 184 192 231
fetal lung	Invitrogen	FLG003	6 15 19 60 89 107 111 113 147 154 158 164-166 190 224 238 242
fetal lung	Clontech	FLG004	99-100
fetal liver-spleen	Columbia University	FLS001	1-7 9 11 17 26 28-29 38-39 41 48 51-53 57-60 72 74 76 84 90-91 93-95 97-102 104-110 112-122 126 132-133 135-136 138 143 149-150 153 159 161 167 172 178 181 191 194 198 200-203 211 213 220 230 238 242 263 265
fetal liver-spleen	Columbia University	FLS002	5-6 9 11 15 18 26 28 32 42 48 51-53 57-60 72 79-80 82-84 89-90 93 95 97-98 101-102 105-110 112-119 126 129 132 134-135 137 153-155 157 164-167 169 172 174 180-181 184 191 194 197 201-202 207 213 220 224 226 230 238 241-242 263 265 268
fetal liver-spleen	Columbia University	FLS003	5 9 21 26 28 90-91 93-94 99-100 106 109-110 113 115-117 121 133 136 143-144 153 164-166 174 178 252
fetal liver	Invitrogen	FLV001	32 35 101-102 106 112 120 126 137 172-173 178 188 240 246
fetal liver	Clontech	FLV002	10 85 89 107 116 120 221 224
fetal liver	Clontech	FLV004	15 58 69-70 81 89-92 104-106 108 111 113-114 122-123 136 147 154-155 164-167 169 172 199 201 203 230 253
fetal muscle	Invitrogen	FMS001	6 14 32 86 107 125 132 154 158 211
fetal muscle	Invitrogen	FMS002	11 14 41 51-53 64 71 74 95 109-110 115 118 129 136 148 178 184 199-200 221 242 253 255
fetal skin	Invitrogen	FSK001	1-4 6 10-11 13 15 24 29 78 86-87 91 97 99-102 105-107 109-110 115 132 134 136 138 147 153-154 158 164-167 169 178 186 188 192 200 210 225 228 234-235 238 240 242
fetal skin	Invitrogen	FSK002	5-6 8 15 28-29 51-53 55 60 71 74 76 78 89 91-92 94 103 105-106 111-112 115 117-118 122-123 136 138-139 144 147 155 157 161 178 188 190 198-201 204 209 211 221 225 230 253 259-260 267 272
umbilical cord	BioChain	FUC001	4-5 28 38-39 78 80-81 84 86 99-102 104-106 109-110 113-116 121 124 126 132-133 138 147 153 158 200 211 216 249 252
fetal brain	GIBCO	HFB001	1-3 8-10 14 16 22 24 26 29 76 78-79 95 101-102 104-105 108 111 113 115 118 125-131 134 162 164-166 172 178 209 220-221 224 244
macrophage	Invitrogen	HMP001	4 41 73 101-102 104 107-108 115 147 154 159 169 183 196-197 199-200 219
infant brain	Columbia University	IB2002	7 10 14 16 22-23 25 29 31 36-39 47 50-53 59-60 64 76 81 87 99-100 105-108 112-113 115 121 135 137-140 146-147 153

119  
Table 1

Tissue origin	Library/RNA source	HYSEQ Library Name	SEQ ID NO:
			158 161-162 167 173 178 192 199 213 224-225 232-234 239-240 242 254 269
infant brain	Columbia University	IB2003	6 11 15-16 29 36-39 47 51-53 64 76 79 87-88 109-110 113 128 132 137 144 146- 147 153-154 158 161-162 173 178 192 199-200 224-225 232 240 242 269
infant brain	Columbia University	IBM002	139 161 242
infant brain	Columbia University	IBS001	10 37 107 109-110 112 162 173 269
lung, fibroblast	Stratagene	LFB001	4-5 15 28 41-42 57 72 76 80 99-100 107 132 153 160 219
lung tumor	Invitrogen	LGT002	1-3 5-6 9-10 21 27-29 32 43 46 48 57 60 78 84 87 104-106 109-113 115 118 122 125 133-134 149 153 159 168 174 177- 178 181 211 214 216 220 235 237-239 242 252 265 267
lymphocytes	ATCC	LPC001	13 41 60 78 84 91 95 99-103 105 107 109- 110 112-113 118 125-126 132-133 143 153 159 173 181 187 200 207 225 240 246 265
leukocyte	GIBCO	LUC001	1-3 5-6 9 11 15 18-19 28 41 43 45 51-53 57 60 74 78 80 82-83 93 95 97 99-100 104-105 107-111 113-115 118 121-123 125-126 132 137 144 146-148 150 155 158-159 178 181 198-199 207 211 213 223 235 246-247 253
leukocyte	Clontech	LUC003	60 99-100 105 132 154
melanoma from-cell-line- ATCC-#CRL- 1424	Clontech	MEL004	99-100 106 120 144 157 169 191 211 219- 220 264
mammary gland	Invitrogen	MMG001	4-7 11 13 15-16 25-26 28 38-39 74 79 84 86-87 90-92 94 101-102 104 106-107 109- 110 112-115 122 129 132 136 138 144 147 153-154 157-158 164-166 168-169 171-172 174-175 178 187-188 192 194 208 221 240 242 263 265
mixture 16 tissues/mRNA	various vendors	SUP002	15 38-39 44 85-86 112 117 120-121 123 126 147 178 186 190 222 224 254 259- 260 272
mixture 16 tissues/mRNA	various vendors	SUP008	99-100 111 114 158 246
mixture 16 tissues/mRNA	various vendors	SUP009	1-3
induced neuron- cells	Stratagene	NTD001	16 29 43 76 79 105 107 132 162
retinoic acid- induced- neuronal-cells	Stratagene	NTR001	47 109-110 115 118 154 157 159 178 199 230
neuronal cells	Stratagene	NTU001	1-3 16 29 60 89 106 109-110 118 143 200 209
pituitary gland	Clontech	PIT004	1-4 51-53 72 77 109-111 113 174 240 247 263 265
placenta	Clontech	PLA003	1-3 30 71 89 97 104 115 161 169 184 199

120  
Table 1

Tissue origin	Library/RNA source	HYSEQ Library Name	SEQ ID NO:
			216
prostate	Clontech	PRT001	10 12 15 18 35 46 80 84 113 121 125 136 154 159 164-166 178 200 211 252 265 267 273
rectum	Invitrogen	REC001	6 32 48 67 80 90 101-102 107 109-110 122 154 159 168 173 192 221 229-230 240 253 265-266
salivary gland	Clontech	SAL001	11 15 35 49 60 84 94 104 109-110 123 134 137 174 178 246
small intestine	Clontech	SIN001	5-6 9 11 13 16 26 28-29 38-39 47 51-53 57 72 76-77 80 86-87 91 93 101-102 104- 105 107 109-110 113-114 120-122 126 132 134 136 144 155 159 164-166 168 181 188 209 234 240 247 252-254 265 267
skeletal muscle	Clontech	SKM001	7 9 14 24 35 42 57 107 109-110 125 150 153 195
spinal cord	Clontech	SPC001	1-3 23-24 38-39 41 46 87 91 99-103 109- 111 113 115 118 125-126 132 145 153 159 161-162 169 181 194 198-200 209 211 224-225 231 247 252 272
adult spleen	Clontech	SPLc01	6 15 82-83 91 107 114 147 159 178 181 202 221 246
stomach	Clontech	STO001	10 15 58 91
thalamus	Clontech	THA002	16 76 87 90 104 132 153 157 162 172 175-176 190 194 211 240
thymus	Clontech	THM001	1-3 26 32 38-39 41 60 107 132 136 157 211 231 246 261 263-264
thymus	Clontech	THMc02	1-3 5 9 15-16 19 21 28 33 38-39 46 51-54 58 71 75 80 82-83 91 93 95 97 103-105 115 122 132-133 147 157 163 173 178 186 190 194 199 204 211 219 225 230 235 246 253 263
thyroid gland	Clontech	THR001	1-7 9 12-13 15 19 28 41 43 45 47 51-52 72 78 80 82-84 86-87 93-95 99-100 104 106- 110 115-116 126 130 136-139 154-155 159-160 163 168 186-187 199-201 210- 212 216 232 242 265 267
trachea	Clontech	TRC001	18 28-29 46 101-102 113 143 149 158 192 194 211 238 240
uterus	Clontech	UTR001	30 38-39 86 121 132 137 150 155
bone marrow	STM001	115	199

\*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
277	gi1321818	Gallus gallus	RING zinc finger protein	1355	91
277	gi2746333	Homo sapiens	RING zinc finger protein (RZF) mRNA, complete cds.	1455	100
277	gi3387925	Homo sapiens	clone 24450 RING zinc finger protein RZF mRNA, complete cds.	1455	100
278	gi2746333	Homo sapiens	RING zinc finger protein (RZF) mRNA, complete cds.	1445	94
278	gi3387925	Homo sapiens	clone 24450 RING zinc finger protein RZF mRNA, complete cds.	1445	94
278	gi14602541	Homo sapiens	ring finger protein 13, clone MGC:13487 IMAGE:3683407, mRNA, complete cds.	1445	94
279	gi2746333	Homo sapiens	RING zinc finger protein (RZF) mRNA, complete cds.	1338	100
279	gi3387925	Homo sapiens	clone 24450 RING zinc finger protein RZF mRNA, complete cds.	1338	100
279	gi14602541	Homo sapiens	ring finger protein 13, clone MGC:13487 IMAGE:3683407, mRNA, complete cds.	1338	100
280	gi10438603	Homo sapiens	cDNA: FLJ22282 fis, clone HRC03861.	1341	96
280	AAB24463	Homo sapiens	Human secreted protein sequence encoded by gene 27 SEQ ID NO:88.	1341	96
280	AAB34813	Homo sapiens	Human secreted protein sequence encoded by gene 41 SEQ ID NO:101.	696	93
281	gi6841548	Homo sapiens	HSPC163	423	100
281	gi12653595	Homo sapiens	HSPC163 protein, clone MGC:772 IMAGE:3163724, mRNA, complete cds.	423	100
281	AAV91543	Homo sapiens	Human secreted protein sequence encoded by gene 93 SEQ ID NO:216.	423	100
282	gi2586350	Homo sapiens	tetraspan (NAG-2) mRNA, complete cds.	842	93
282	gi2997747	Homo sapiens	tetraspan TM4SF (TSPAN-4) mRNA, complete cds.	842	93
282	gi12653241	Homo sapiens	transmembrane 4 superfamily member 7, clone MGC:8437 IMAGE:2821236, mRNA, complete cds.	842	93
283	gi15080477	Homo sapiens	Similar to RIKEN cDNA 2310010G13 gene, clone MGC:9810 IMAGE:3860434, mRNA, complete cds.	2037	97
283	gi9104959	Xylella fastidiosa 9a5c	beta-lactamase induction signal transducer protein	161	29
283	gi1778812	Neisseria gonorrhoeae	No definition line found	259	27
284	gi12053215	Homo sapiens	mRNA; cDNA DKFZp434K2435 (from clone DKFZp434K2435); complete cds.	2762	100
284	AAV87197	Homo sapiens	Human secreted protein sequence SEQ ID NO:236.	86	24
284	AAV27598	Homo sapiens	Human secreted protein encoded by gene No. 32.	63	29
285	gi10438815	Homo sapiens	cDNA: FLJ22427 fis, clone HRC09013.	4487	98
285	gi15076843	Homo sapiens	pecanex-like protein 1 mRNA, complete cds.	759	44
285	gi13171105	Takifugu rubripes	pecanex	685	44
286	gi2828808	Bacillus subtilis	glucose transporter	100	23
286	gi14023148	Mesorhizobiu	probable fosmidomycin resistance protein	112	25

122  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		m loti			
286	gi2650264	Archaeoglobus fulgidus	oxalate/formate antiporter (oxlT-2)	102	23
287	gi180137	Homo sapiens	Human membrane cofactor protein (MCP) mRNA, complete cds.	1980	96
287	AAW27484	Homo sapiens	Human MCP.	1980	96
287	gi512457	Homo sapiens	membrane cofactor protein	1976	95
288	gi10437579	Homo sapiens	cDNA: FLJ21472 fis, clone COL04936.	1019	100
288	AAE01687	Homo sapiens	Human gene 16 encoded secreted protein HDPMM88, SEQ ID NO:99.	1019	100
288	gi14043759	Homo sapiens	clone IMAGE:4111596, mRNA, partial cds.	563	58
289	AAAY41401	Homo sapiens	Human secreted protein encoded by gene 94 clone HLYCH68.	392	100
289	AAB08863	Homo sapiens	Amino acid sequence of a human secretory protein.	392	100
289	gi575398	Saccharomyces cerevisiae	regulator of carbon catabolite repression	54	57
290	gi14250010	Homo sapiens	clone MGC:14489 IMAGE:4244549, mRNA, complete cds.	2035	99
290	gi1495419	Homo sapiens	H.sapiens ART3 gene.	1713	97
290	gi2677616	Mus musculus	NAD(P)(+)-arginine ADP-ribosyltransferase	1080	58
291	gi13182757	Homo sapiens	HTPAP mRNA, complete cds.	598	100
291	AAB70690	Homo sapiens	Human hDPP protein sequence SEQ ID NO:7.	598	100
291	gi14020949	Arabidopsis thaliana	phosphatidic acid phosphatase	250	38
292	AAB88418	Homo sapiens	Human membrane or secretory protein clone PSEC0181.	725	100
292	gi2909844	Homo sapiens	prostate stem cell antigen (PSCA) mRNA, complete cds.	109	32
292	gi9367212	Homo sapiens	mRNA for prostate stem cell antigen (PSCA gene).	109	32
293	gi12718841	Mus musculus	Skullin	283	38
293	gi4191356	Mus musculus	claudin-6	281	38
293	gi13543081	Mus musculus	claudin 6	281	38
294	gi2618609	Capra hircus	mhc class II DRA	636	80
294	gi165868	Ovis aries	MHC Ovar-DR-alpha	632	79
294	gi207708	Sciurus aberti	MHC class II DR-alpha	652	82
295	gi14025214	Mesorhizobium loti	probable amidase	348	31
295	gi7226601	Neisseria meningitidis MC58	Glu-tRNA(Gln) amidotransferase, subunit A	398	28
295	gi7380209	Neisseria meningitidis Z2491	Glu-tRNA(Gln) amidotransferase subunit A	387	27
296	gi12620132	Homo sapiens	renal sodium/sulfate cotransporter mRNA, complete cds.	3100	100
296	gi10439272	Homo sapiens	cDNA: FLJ22760 fis, clone KAI1A0881.	3096	99
296	gi310183	Rattus norvegicus	sodium dependent sulfate transporter	2627	82
297	gi12653037	Homo sapiens	clone IMAGE:3355813, mRNA, partial	1574	100

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			cds.		
297	AAAY44245	Homo sapiens	Human cell signalling protein-8.	1208	100
297	AAW64220	Homo sapiens	Human secreted protein from clone CG300 3.	1195	98
298	gi9588085	Homo sapiens	mRNA for TAPL, complete cds.	2338	99
298	gi9622987	Homo sapiens	ATP-binding cassette protein ABCB9 (ABCB9) mRNA, complete cds.	2338	99
298	AAE02437	Homo sapiens	Human ATP binding cassette, ABCB9 transporter protein.	2338	99
299	AAAY87237	Homo sapiens	Human signal peptide containing protein HSPP-14 SEQ ID NO:14.	110	30
299	AAB87384	Homo sapiens	Human gene 43 encoded secreted protein HSLGM81, SEQ ID NO:125.	110	30
299	AAB87410	Homo sapiens	Human gene 43 encoded secreted protein HSYBM41, SEQ ID NO:151.	110	30
300	gi3874886	Caenorhabditis elegans	C41C4.2	557	49
300	gi13785612	Mus musculus	sideroflexin I	404	39
300	gi13543138	Mus musculus	RIKEN cDNA 2810002O05 gene	404	39
301	gi5114275	Homo sapiens	MAB21L2 (MAB21L2) gene, complete cds.	113	33
301	gi9964007	Homo sapiens	MAB21L2 protein (MAB21L2) mRNA, complete cds.	113	33
301	gi14134002	Homo sapiens	MAB21L2 protein mRNA, complete cds.	113	33
302	gi7020704	Homo sapiens	cDNA FLJ20533 fis, clone KAT10931.	829	98
302	gi15030135	Mus musculus	RIKEN cDNA 1110020A09 gene	777	60
302	gi5824484	Caenorhabditis elegans	F32D8.5b	111	25
303	gi10433539	Homo sapiens	cDNA FLJ12133 fis, clone MAMMA1000278.	319	30
303	AAB93897	Homo sapiens	Human protein sequence SEQ ID NO:13844.	319	30
303	AAW64461	Homo sapiens	Human secreted protein from clone B121.	313	30
304	gi6841548	Homo sapiens	HSPC163	489	100
304	gi12653595	Homo sapiens	HSPC163 protein, clone MGC:772 IMAGE:3163724, mRNA, complete cds.	489	100
304	AAAY91543	Homo sapiens	Human secreted protein sequence encoded by gene 93 SEQ ID NO:216.	489	100
305	gi4877582	Homo sapiens	lipoma HMGIC fusion partner (LHFP) mRNA, complete cds.	222	28
305	AAAY87336	Homo sapiens	Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	222	28
305	AAW88508	Homo sapiens	Human stomach cancer clone HP10480-encoded membrane protein.	94	26
306	AAB87576	Homo sapiens	Human PRO3579.	1125	98
306	gi2315510	Caenorhabditis elegans	similar to 1-acyl-glycerol-3-phosphate acyltransferases	501	45
306	gi3877657	Caenorhabditis elegans	contains similarity to Pfam domain: PF01553 (Acyltransferase), Score=144.3, E-value=7.1e-40, N=1	364	44
307	AAAY94954	Homo sapiens	Human secreted protein clone iw66_1 protein sequence SEQ ID NO:114.	596	68
307	gi7259234	Mus musculus	contains transmembrane (TM) region	562	63
307	AAB62810	Homo sapiens	Human nervous system associated protein	536	60



Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NSPRT3 amino acid sequence.		
308	gi4580997	Mus musculus	cAMP inducible 2 protein	2377	87
308	gi7543982	Homo sapiens	mRNA for glycerol 3-phosphate permease (SLC37A1 gene).	842	60
308	gi11095363	Homo sapiens	glycerol 3-phosphate permease (SLC37A1) mRNA, complete cds.	836	60
309	AAG71797	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1478.	755	100
309	gi12007408	Mus musculus	B1 olfactory receptor	625	79
309	gi12007420	Mus musculus	B5 olfactory receptor	609	82
310	gi12803871	Homo sapiens	clone MGC:4170 IMAGE:3618204, mRNA, complete cds.	373	100
310	gi3881055	Caenorhabditis elegans	Y48A6B.1	57	59
310	gi13398356	Trichoplusia ni	acyl-CoA delta-11 desaturase	46	53
311	gi11128456	Homo sapiens	nicotinic acetylcholine receptor subunit alpha 10 mRNA, complete cds.	2370	100
311	gi13173184	Homo sapiens	nicotinic acetylcholine receptor subunit alpha 10 (CHRNA10) gene, complete cds.	2370	100
311	gi12053839	Homo sapiens	mRNA for neuronal nicotinic acetylcholine alpha 10 subunit (NACHRA10 gene).	2370	100
312	gi14328885	Mus musculus	spermatogenic immunoglobulin superfamily protein	630	40
312	gi7767239	Homo sapiens	nectin-like protein 2 (NECL2) mRNA, complete cds.	628	41
312	gi4519602	Homo sapiens	IGSF4 gene, exon 10 and complete cds.	625	40
313	AAA40083.aa1	Homo sapiens	Human brain-specific transmembrane glycoprotein encoding cDNA.	1637	54
313	AAB09968	Homo sapiens	Human brain-specific transmembrane glycoprotein.	1637	54
313	AAB12448	Homo sapiens	Human hh00149 protein SEQ ID NO:4.	1637	54
314	gi14017379	Homo sapiens	tumor endothelial marker 7 precursor (TEM7) mRNA, complete cds.	2691	100
314	AAB31211	Homo sapiens	Amino acid sequence of human polypeptide PRO6003.	1297	57
314	AAW58986	Homo sapiens	Homo sapiens adult brain clone CC194_4 encoded protein.	560	99
315	gi14017379	Homo sapiens	tumor endothelial marker 7 precursor (TEM7) mRNA, complete cds.	2592	97
315	AAB31211	Homo sapiens	Amino acid sequence of human polypeptide PRO6003.	1040	53
315	AAW58986	Homo sapiens	Homo sapiens adult brain clone CC194_4 encoded protein.	461	87
316	AAG71567	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1248.	1414	100
316	AAG71576	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1257.	726	52
316	AAG72477	Homo sapiens	Human OR-like polypeptide query sequence, SEQ ID NO: 2158.	726	52
317	gi14495648	Homo sapiens	clone MGC:15606 IMAGE:3163718, mRNA, complete cds.	2958	100
317	AAB74709	Homo sapiens	Human membrane associated protein MEMAP-15.	338	31

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
317	gi7020023	Homo sapiens	cDNA FLJ20127 fis, clone COL06176.	149	29
318	AAB88430	Homo sapiens	Human membrane or secretory protein clone PSEC0205.	2226	99
318	AAY44363	Homo sapiens	Human cell cycle regulation protein-4.	1827	100
318	AAB08956	Homo sapiens	Human secreted protein sequence encoded by gene 24 SEQ ID NO:113.	1819	99
319	AAY19506	Homo sapiens	Amino acid sequence of a human secreted protein.	1120	100
319	gi11177546	Homo sapiens	LIM2 (LIM2) and natural killer group 7 (NKG7) genes, complete cds.	90	26
319	gi13445660	Homo sapiens	MP19 (LIM2) mRNA, complete cds, alternatively spliced.	90	26
320	gi784990	Homo sapiens	H.sapiens DNA for 5-HT5A exon1.	1645	100
320	gi6064324	unidentified	GENE DU RECEPTEUR 5HT5A HUMAIN	1611	98
320	AAR45848	Homo sapiens	Human 5HT5a serotonin receptor.	1611	98
321	gi2695874	Homo sapiens	H.sapiens mRNA for P2Y-like G-protein coupled receptor.	175	28
321	AAR53752	Homo sapiens	Seven transmembrane receptor (R12).	175	28
321	AAW07617	Homo sapiens	Human G-protein thrombin-like receptor.	175	28
322	AAY25806	Homo sapiens	Human secreted protein fragment encoded from gene 23.	1663	98
322	gi5901846	Drosophila melanogaster	BcDNA.GH12144	627	43
322	AAB12140	Homo sapiens	Hydrophobic domain protein isolated from WERI-RB cells.	353	36
323	gi10438949	Homo sapiens	cDNA: FLJ22529 fis, clone HRC12842.	1290	100
323	AAB12119	Homo sapiens	Hydrophobic domain protein from clone HP02869 isolated from KB cells.	448	100
323	gi13384443	Caenorhabditis elegans	similar to 1-acyl-glycerol-3-phosphate acyltransferases	294	26
324	AAY25736	Homo sapiens	Human secreted protein encoded from gene 26.	343	100
324	gi14530705	Caenorhabditis elegans	Similarity to C.elegans UNC-7 protein (SW:UNC7 CAEEL), contains similarity to Pfam domain: PF00876 (Innexin), Score=640.8, E-value=2.4e-189, N=1	75	36
324	gi142083	Anabaena sp.	ribulose 1,5-bisphosphate carboxylase/oxygenase small subunit	63	41
325	AAB44336	Homo sapiens	Human secreted protein encoded by gene 2 clone HROAM11.	169	100
325	AAG03801	Homo sapiens	Human secreted protein, SEQ ID NO: 7882.	64	41
325	gi6139004	Echinococcus multilocularis	NADH dehydrogenase subunit 6	45	55
326	gi10566471	Mus musculus	Gliacolin	1284	94
326	gi14278927	Mus musculus	gliacolin	1284	94
326	gi3747097	Homo sapiens	Clq-related factor mRNA, complete cds.	974	71
327	gi13506225	Mus musculus	ST7 protein form1 splice variant a	2996	99
327	gi9230665	Homo sapiens	FAM4A1 splice variant a (FAM4A1) mRNA, complete cds.	1761	96
327	gi13506227	Mus musculus	ST7 protein form1 splice variant b	1761	96
328	gi9230665	Homo sapiens	FAM4A1 splice variant a (FAM4A1) mRNA, complete cds.	2496	97

126  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
328	gi13506227	Mus musculus	ST7 protein form1 splice variant b	2489	96
328	gi13506225	Mus musculus	ST7 protein form1 splice variant a	1366	92
329	gi9230667	Homo sapiens	FAM4A1 splice variant b (FAM4A1) mRNA, complete cds.	2862	97
329	gi13506225	Mus musculus	ST7 protein form1 splice variant a	2848	96
329	gi9230665	Homo sapiens	FAM4A1 splice variant a (FAM4A1) mRNA, complete cds.	1608	92
330	gi292057	Homo sapiens	Human EBV induced G-protein coupled receptor (EBI2) mRNA, complete cds.	321	38
330	AAR54080	Homo sapiens	Epstein Barr virus induced (EBI-2) polypeptide.	321	38
330	AAW53623	Homo sapiens	Epstein Barr virus induced gene 2 (EBI-2).	321	38
331	gi10434308	Homo sapiens	cDNA FLJ12672 fis, clone NT2RM4002339.	3584	99
331	AAB94231	Homo sapiens	Human protein sequence SEQ ID NO:14604.	3584	99
331	gi10436632	Homo sapiens	cDNA FLJ14225 fis, clone NT2RP3004051.	3570	100
332	gi3462455	Mus musculus	junctional adhesion molecule	116	28
332	AAV23325	Homo sapiens	A33 related antigen JAM.	116	28
332	gi8650528	Rattus norvegicus	junctional adhesion molecule JAM	109	27
333	gi14250676	Homo sapiens	Similar to RIKEN cDNA 2310002F18 gene, clone MGC:10413 IMAGE:3954787, mRNA, complete cds.	1977	99
333	AAV27589	Homo sapiens	Human secreted protein encoded by gene No. 23.	1578	100
333	gi12082328	Arabidopsis thaliana	para-hydroxy benzoate polyprenyl diphosphate transferase	792	64
334	gi12655071	Homo sapiens	transmembrane 4 superfamily member 4, clone MGC:1477 IMAGE:3051146, mRNA, complete cds.	859	98
334	gi953239	Homo sapiens	Human intestinal and liver tetraspan membrane protein (il-TMP) mRNA, complete cds.	859	98
334	gi11493837	Rattus norvegicus	tetraspan protein LRTM4	791	85
336	gi14336694	Homo sapiens	16p13.3 sequence section 2 of 8.	4100	99
336	gi10716072	Homo sapiens	mRNA for M83 protein, complete cds.	4089	99
336	gi10716074	Mus musculus	M83 protein	3115	75
337	gi11023146	Homo sapiens	corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) mRNA, complete cds.	2056	100
337	gi11023149	Homo sapiens	intestinal N-acetylglucosamine-6-O-sulfotransferase (CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) genes, complete cds.	2056	100
337	gi12060804	Homo sapiens	N-acetylglucosamine 6-O-sulfotransferase GST-4beta mRNA, complete cds.	2056	100
338	AAG71850	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1531.	1142	71
338	AAG71809	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1490.	1049	74
338	AAG71818	Homo sapiens	Human olfactory receptor polypeptide,	1014	68

127  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SEQ ID NO: 1499.		
339	AAG71850	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1531.	1128	71
339	AAG71809	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1490.	1035	74
339	AAG71818	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1499.	1014	68
340	gi7960136	Homo sapiens	neurologin 3 isoform gene, complete cds, alternatively spliced.	4557	100
340	gi1145791	Rattus norvegicus	neurologin 3	4505	98
340	gi7960135	Homo sapiens	neurologin 3 isoform gene, complete cds, alternatively spliced.	3623	96
341	gi5525078	Rattus norvegicus	seven transmembrane receptor	788	31
341	AAAY57288	Homo sapiens	Human GPCR protein (HGPRP) sequence (clone ID 3036563).	752	29
341	AAAY40440	Homo sapiens	Human brain-derived G-protein coupled receptor protein.	746	29
342	AAG71424	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1105.	853	88
342	AAG72315	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1996.	915	96
342	AAG71431	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1112.	595	60
343	gi10434098	Homo sapiens	cDNA FLJ12547 fis, clone NT2RM4000634.	1612	84
343	AAB95124	Homo sapiens	Human protein sequence SEQ ID NO:17122.	1612	84
343	gi854065	Human herpesvirus 6	U88	809	52
344	AAG71823	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1504.	1627	100
344	AAG71859	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1540.	1085	67
344	AAG72185	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1866.	980	60
345	AAAY91625	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	1968	94
345	AAU00437	Homo sapiens	Human dendritic cell membrane protein FIRE.	1925	78
345	AAAY59300	Homo sapiens	Human EGPCR polypeptide.	1174	57
346	AAAY91625	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	1968	94
346	AAU00437	Homo sapiens	Human dendritic cell membrane protein FIRE.	1925	78
346	AAAY59300	Homo sapiens	Human EGPCR polypeptide.	1174	57
347	gi4098462	Sus scrofa	luteinizing hormone beta subunit	41	53
347	gi12232003	Cercopithecus aethiops	NADH dehydrogenase subunit 5	81	32
348	AAW74874	Homo sapiens	Human secreted protein encoded by gene 146 clone HSNK17.	349	100
348	gi3329179	Chlamydia trachomatis	Phosphoglycerate Mutase	68	33

128  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
348	gi9105100	Xylella fastidiosa 9a5c	transport protein	68	46
349	AA04301	Homo sapiens	Human secreted protein encoded by gene 9.	82	33
349	gi15004512	Podophyllum peltatum	succinate dehydrogenase subunit 3	79	32
349	gi841378	Saccharomyces cerevisiae	Gpi2p	90	34
350	AAB88406	Homo sapiens	Human membrane or secretory protein clone PSEC0162.	1421	99
350	AAW88579	Homo sapiens	Secreted protein encoded by gene 46 clone HCFMV39.	479	95
350	AA041111	Homo sapiens	Human TANGO 129 (T129) mature protein.	225	35
351	gi292793	Homo sapiens	(clone HBVT72) T cell receptor beta chain (TCRB) mRNA, VDJC region, partial cds.	636	98
351	gi457274	Homo sapiens	Human T-cell receptor beta chain gene, V region, partial cds.	479	98
351	gi495428	Macaca mulatta	T cell receptor beta chain	477	85
352	AA010839	Homo sapiens	Amino acid sequence of a human secreted protein.	225	95
352	gi15163613	Agrobacterium tumefaciens	AGR_pTi_226p	66	40
352	gi903711	Daucus carota	cytochrome oxidase II	59	36
353	AA016784	Homo sapiens	Human secreted protein (clone co1000_1).	488	100
353	gi1850866	Macropus robustus	ATPase subunit 8	68	31
353	AA041439	Homo sapiens	Fragment of human secreted protein encoded by gene 24.	63	43
354	gi6573749	Arabidopsis thaliana	F20B24.9	58	38
354	gi325236	Influenza B virus	nb	61	34
354	AAR11254	Homo sapiens	Human IL-4 receptor.	60	52
355	gi12652903	Homo sapiens	clone MGC:3103 IMAGE:3350518, mRNA, complete cds.	1704	100
355	AAA40083aal	Homo sapiens	Human brain-specific transmembrane glycoprotein encoding cDNA.	1019	43
355	AAB09968	Homo sapiens	Human brain-specific transmembrane glycoprotein.	1019	43
356	gi10439087	Homo sapiens	cDNA: FLJ22625 fis, clone HSI06009.	1792	100
356	AA041389	Homo sapiens	Human secreted protein encoded by gene 82 clone HOUH51.	1555	94
356	AA041747	Homo sapiens	Human PRO534 protein sequence.	1555	94
358	gi13676372	Homo sapiens	clone MGC:4595 IMAGE:3345729, mRNA, complete cds.	1886	98
358	AA041690	Homo sapiens	Human PRO329 protein sequence.	1886	98
358	AAB44246	Homo sapiens	Human PRO329 (UNQ291) protein sequence SEQ ID NO:45.	1886	98
359	gi13676372	Homo sapiens	clone MGC:4595 IMAGE:3345729, mRNA, complete cds.	1905	99
359	AA041690	Homo sapiens	Human PRO329 protein sequence.	1905	99
359	AAB44246	Homo sapiens	Human PRO329 (UNQ291) protein	1905	99

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			sequence SEQ ID NO:45.		
360	AAW74807	Homo sapiens	Human secreted protein encoded by gene 79 clone HSKNE46.	270	100
360	gi2145070	Mus musculus	m17r splice variant	49	46
360	AAB34697	Homo sapiens	Human secreted protein encoded by DNA clone vq6 1.	66	45
361	gi6959684	Mus musculus	glycolipid transfer protein	103	26
361	gi14041214	Human herpesvirus 4	EBNA-LP protein	76	36
361	gi6959686	Homo sapiens	glycolipid transfer protein mRNA, complete cds.	93	24
362	gi13623231	Homo sapiens	Similar to RIKEN cDNA 1200013A08 gene, clone MGC:3047 IMAGE:3343261, mRNA, complete cds.	2337	100
362	gi14041843	Homo sapiens	cDNA FLJ14363 fis, clone HEMBA1000719.	2270	98
362	AAB92464	Homo sapiens	Human protein sequence SEQ ID NO:10520.	2270	98
363	gi10438446	Homo sapiens	cDNA: FLJ22167 fis, clone HRC00584.	1644	100
364	gi12053067	Homo sapiens	mRNA; cDNA DKFZp434I2117 (from clone DKFZp434I2117).	1237	100
364	gi10438603	Homo sapiens	cDNA: FLJ22282 fis, clone HRC03861.	649	48
364	AAB24463	Homo sapiens	Human secreted protein sequence encoded by gene 27 SEQ ID NO:88.	649	48
365	gi12483888	Homo sapiens	solute carrier 19A3 mRNA, complete cds.	2549	100
365	gi14582572	Homo sapiens	orphan transporter SLC19A3 (SLC19A3) mRNA, complete cds.	2549	100
365	gi12483890	Mus musculus	solute carrier 19A3	1716	68
366	AAB74721	Homo sapiens	Human membrane associated protein MEMAP-27.	558	100
366	AAG03412	Homo sapiens	Human secreted protein, SEQ ID NO: 7493.	464	100
366	gi4929751	Homo sapiens	CGI-141 protein mRNA, complete cds.	406	55
367	gi10434145	Homo sapiens	cDNA FLJ12576 fis, clone NT2RM4001032.	2598	100
367	gi12803561	Homo sapiens	clone MGC:2991 IMAGE:3160297, mRNA, complete cds.	2598	100
367	AAB94138	Homo sapiens	Human protein sequence SEQ ID NO:14406.	2598	100
368	gi4519535	Homo sapiens	CYP4F2 gene for leukotriene B4 omega hydroxylase, exon 13.	1227	65
368	gi1857022	Homo sapiens	Human mRNA for leukotriene B4 omega-hydroxylase, complete cds.	1227	65
368	gi10303605	Homo sapiens	CYP4F11 mRNA, complete cds.	1219	64
369	gi10438815	Homo sapiens	cDNA: FLJ22427 fis, clone HRC09013.	4518	100
369	gi15076843	Homo sapiens	pecanex-like protein 1 mRNA, complete cds.	762	44
369	gi13171105	Takifugu rubripes	pecanex	578	42
370	gi12656635	Homo sapiens	transmembrane gamma-carboxylglutamic acid protein 4 TMG4 mRNA, complete cds.	1201	100
370	gi14603178	Homo sapiens	transmembrane gamma-carboxylglutamic acid protein 4, clone MGC:19793	1201	100

130  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			IMAGE:3841745, mRNA, complete cds.		
370	AAB61219	Homo sapiens	Human TANGO 292 protein.	1201	100
371	gi7689031	Homo sapiens	uncharacterized hypothalamus protein HARP11 mRNA, complete cds.	1847	100
371	gi15080516	Homo sapiens	Similar to uncharacterized hypothalamus protein HARP11, clone MGC:9273 IMAGE:3862712, mRNA, complete cds.	1847	100
371	AAY53029	Homo sapiens	Human secreted protein clone cw1640_1 protein sequence SEQ ID NO:64.	1847	100
372	gi10440079	Homo sapiens	cDNA: FLJ23403 fis, clone HEP18857.	2817	100
372	AAY53635	Homo sapiens	A bone marrow secreted protein designated BMS53.	758	50
372	gi10439735	Homo sapiens	cDNA: FLJ23144 fis, clone LNG09262.	771	100
373	gi7023450	Homo sapiens	cDNA FLJ11036 fis, clone PLACE1004289.	980	87
373	AAB93444	Homo sapiens	Human protein sequence SEQ ID NO:12686.	980	87
373	gi1199697	Athalia rosae	vitellogenin	107	42
374	gi13447851	Macaca mulatta	killer immunoglobulin-like receptor KIR3DL7	77	31
374	gi190203	Homo sapiens	Human cardiac potassium channel (KCNA5) mRNA, complete cds.	83	33
374	gi308765	Homo sapiens	Human voltage-gated potassium channel (HK2) mRNA, complete cds.	82	35
375	gi5542014	Homo sapiens	DKC1 gene, exons 1 to 11.	1574	99
375	gi3873221	Homo sapiens	dyskerin (DKC1) mRNA, complete cds.	1574	99
375	gi14603090	Homo sapiens	dyskeratosis congenita 1, dyskerin, clone MGC:15313 IMAGE:4303933, mRNA, complete cds.	1574	99
376	gi5542014	Homo sapiens	DKC1 gene, exons 1 to 11.	2399	95
376	gi3873221	Homo sapiens	dyskerin (DKC1) mRNA, complete cds.	2326	94
376	gi14603090	Homo sapiens	dyskeratosis congenita 1, dyskerin, clone MGC:15313 IMAGE:4303933, mRNA, complete cds.	2326	94
377	gi12653555	Homo sapiens	lysophospholipase-like, clone MGC:1216 IMAGE:3163689, mRNA, complete cds.	907	100
377	gi13623261	Homo sapiens	lysophospholipase-like, clone MGC:10338 IMAGE:3945191, mRNA, complete cds.	907	100
377	gi1763011	Homo sapiens	Human lysophospholipase homolog (HU-K5) mRNA, complete cds.	907	100
378	gi12653555	Homo sapiens	lysophospholipase-like, clone MGC:1216 IMAGE:3163689, mRNA, complete cds.	903	100
378	gi13623261	Homo sapiens	lysophospholipase-like, clone MGC:10338 IMAGE:3945191, mRNA, complete cds.	903	100
378	gi1763011	Homo sapiens	Human lysophospholipase homolog (HU-K5) mRNA, complete cds.	903	100
379	AAY94946	Homo sapiens	Human secreted protein clone cd205_2 protein sequence SEQ ID NO:98.	571	93
379	AAY53051	Homo sapiens	Human secreted protein clone dd119_4 protein sequence SEQ ID NO:108.	324	63
379	gi4097381	Heteractis magnifica	potassium channel toxin HmK	61	41

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
380	gi6523817	Homo sapiens	S1R protein (S1R) mRNA, complete cds.	928	93
380	gi4929707	Homo sapiens	CGI-119 protein mRNA, complete cds.	928	93
380	AAY77122	Homo sapiens	Human neurotransmission-associated protein (NTAP) 414692.	928	93
381	gi6739575	Mus musculus	TBX2 protein	696	80
381	gi6980032	Mus musculus	ARL-6 interacting protein-1	696	80
381	AAB54057	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:509.	70	28
382	gi13432057	Homo sapiens	NYD-TSPG mRNA, complete cds.	206	25
382	AAB95759	Homo sapiens	Human protein sequence SEQ ID NO:18680.	142	29
382	gi14550463	Homo sapiens	DKFZP434B103 protein, clone MGC:15207 IMAGE:3841498, mRNA, complete cds.	106	32
383	AAY48312	Homo sapiens	Human prostate cancer-associated protein 9.	1509	100
383	gi12654077	Homo sapiens	clone IMAGE:3458173, mRNA, partial cds.	1191	100
383	AAY73387	Homo sapiens	HTRM clone 3340290 protein sequence.	763	82
384	gi14042559	Homo sapiens	cDNA FLJ14784 fis, clone NT2RP4000713.	2492	100
384	AAB93185	Homo sapiens	Human protein sequence SEQ ID NO:12134.	2492	100
384	AAB56514	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1092.	765	98
385	gi12044473	Homo sapiens	mRNA; cDNA DKFZp761D0211 (from clone DKFZp761D0211).	2875	100
385	gi14336686	Homo sapiens	16p13.3 sequence section 1 of 8.	2786	98
385	AAB58984	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 692.	759	94
386	gi14336686	Homo sapiens	16p13.3 sequence section 1 of 8.	2811	100
386	gi12044473	Homo sapiens	mRNA; cDNA DKFZp761D0211 (from clone DKFZp761D0211).	2799	98
386	AAB58984	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 692.	683	89
387	gi3879783	Caenorhabditis elegans	Similarity to Salmonella regulatory protein UHPC (SW:UHPC SALTY)	281	53
387	gi7268507	Arabidopsis thaliana	glycerol-3-phosphate permease like protein	207	44
387	AAB39202	Homo sapiens	Human secreted protein sequence encoded by gene 24 SEQ ID NO:82.	194	38
388	gi14860862	Homo sapiens	polyamine oxidase isoform-1 mRNA, complete cds.	638	52
388	gi7021037	Homo sapiens	cDNA FLJ20746 fis, clone HEP06040.	637	52
388	AAB12164	Homo sapiens	Hydrophobic domain protein from clone HP10673 isolated from Thymus cells.	637	52
389	gi5911897	Homo sapiens	mRNA; cDNA DKFZp586B1417 (from clone DKFZp586B1417); partial cds.	6467	96
389	gi14424668	Homo sapiens	clone MGC:14927 IMAGE:4298580, mRNA, complete cds.	4267	94
389	gi10438036	Homo sapiens	cDNA: FLJ21846 fis, clone HEP01887.	4259	94
390	gi13529623	Mus musculus	Similar to RIKEN cDNA 4930418P06 gene	1408	81
390	gi5656743	Homo sapiens	BAC clone CTB-122E10 from 7q11.23-	105	25



132  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			q21.1, complete sequence.		
390	AAB58323	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 661.	105	25
391	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene, clone MGC:19636 IMAGE:2822323, mRNA, complete cds.	754	96
391	AAB36613	Homo sapiens	Human FLEXHT-35 protein sequence SEQ ID NO:35.	754	96
391	gi7022832	Homo sapiens	cDNA FLJ10661 fis, clone NT2RP2006106.	240	90
392	gi10439204	Homo sapiens	cDNA: FLJ22709 fis, clone HSI13338.	304	39
392	AAB56085	Homo sapiens	Human secreted protein sequence encoded by gene 9 SEQ ID NO:179.	304	39
392	gi7407643	Canis familiaris	occludin 1B	177	32
393	AAB18993	Homo sapiens	Amino acid sequence of a human transmembrane protein.	1212	70
393	gi15079979	Homo sapiens	Similar to RIKEN cDNA 3830408P04 gene, clone MGC:19609 IMAGE:3640970, mRNA, complete cds.	1211	70
393	gi13111831	Homo sapiens	clone IMAGE:3451448, mRNA, partial cds.	980	68
394	AAAY59713	Homo sapiens	Secreted protein 76-20-3-H1-FL1.	865	92
394	gi4220892	Homo sapiens	transcriptional co-activator CRSP34 (CRSP34) mRNA, complete cds.	920	95
394	gi7141322	Homo sapiens	p37 TRAP/SMCC/PC2 subunit mRNA, complete cds.	919	95
395	gi3880799	Caenorhabditis elegans	Y39A1B.2	837	33
395	gi1707052	Caenorhabditis elegans	similar to drosophila and mouse patched proteins	616	35
395	gi861251	Caenorhabditis elegans	weakly similar to C. elegans protein F54G8.5 and to C. elegans protein F44F4.4	475	31
396	gi765240	human, liver, mRNA, 1731 nt]. [Homo sapiens	hPPAR alpha =peroxisome proliferator activated receptor alpha	2011	99
396	AAR74053	Homo sapiens	Human peroxisome proliferator activated receptor.	2011	99
396	AAB20342	Homo sapiens	Peroxisome proliferator-activated receptor alpha.	2011	99
397	AAB43983	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1428.	1692	100
397	AAA88691 aa1	Homo sapiens	Human transmembrane protein NPCAHH01 cDNA.	1410	100
397	gi5565977	Homo sapiens	transmembrane protein BRI (BRJ) mRNA, complete cds.	1409	100
398	gi4894991	Drosophila melanogaster	sodium-hydrogen exchanger NHE1	1362	61
398	gi3979941	Caenorhabditis elegans	contains similarity to Pfam domain: PF00999 (Sodium/hydrogen exchanger family), Score=354.0, E-value=5.3e-103, N=1	1059	46

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
398	gi14150471	Homo sapiens	nonselective sodium potassium/proton exchanger (NHE7) mRNA, complete cds.	679	40
399	gi7023154	Homo sapiens	cDNA FLJ10856 fis, clone NT2RP4001547.	1617	99
399	AAAY28810	Homo sapiens	nn296_2 secreted protein.	1617	99
399	AAB93258	Homo sapiens	Human protein sequence SEQ ID NO:12282.	1617	99
400	AAG00388	Homo sapiens	Human secreted protein, SEQ ID NO: 4469.	316	100
400	gi11967794	Echinops telfairi	NADH dehydrogenase subunit 4L	61	29
400	gi3211979	Homo sapiens	sarco-/endoplasmic reticulum Ca-ATPase 3 (ATP2A3) mRNA, alternatively spliced, partial cds.	54	39
401	gi14043649	Homo sapiens	clone MGC:14161 IMAGE:4111078, mRNA, complete cds.	253	33
401	gi2623016	Methanothermobacter thermautotrophicus	heterodisulfide reductase, subunit C	88	30
401	gi4262178	Arabidopsis thaliana	25726	87	28
402	gi6164616	Homo sapiens	F-box protein Fbl3b (FBL3B) mRNA, partial cds.	128	26
402	AAAY83075	Homo sapiens	F-box protein FBP-3b.	128	26
402	AAAY83043	Homo sapiens	F-box protein FBP-3.	109	23
403	AAB98207	Homo sapiens	Human P24 protein-22 SEQ ID NO:2.	1009	99
403	gi1890141	Mus musculus	P24 protein	940	91
403	gi10439977	Homo sapiens	cDNA: FLJ23329 fis, clone HEP12646.	274	38
404	gi13276693	Homo sapiens	mRNA; cDNA DKFZp761F069 (from clone DKFZp761F069); complete cds.	807	70
404	gi7020303	Homo sapiens	cDNA FLJ20300 fis, clone HEP06465.	539	39
404	AAB67575	Homo sapiens	Amino acid sequence of a human hydrolytic enzyme HYENZ7.	435	33
405	gi3878748	Caenorhabditis elegans	M176.4	98	24
405	gi7542459	Taeniopygia guttata	SWS1 opsin	92	29
405	AAB76874	Homo sapiens	Human lung tumour protein related protein sequence SEQ ID NO:799.	65	51
406	gi3880799	Caenorhabditis elegans	Y39A1B.2	634	25
406	gi861251	Caenorhabditis elegans	weakly similar to C. elegans protein F54G8.5 and to C. elegans protein F44F4.4	261	24
406	gi1255388	Caenorhabditis elegans	similar to drosophila membrane protein PATCHED (SP: P18502)	255	26
407	gi14603058	Homo sapiens	clone IMAGE:4134852, mRNA, partial cds.	1067	100
407	gi1016178	Cyanophora paradoxa	PsaE	53	32
407	gi12724543	Lactococcus lactis subsp. lactis	UNKNOWN PROTEIN	78	43

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
408	AAB12150	Homo sapiens	Hydrophobic domain protein isolated from HT-1080 cells.	952	100
408	gi13096862	Mus musculus	RIKEN cDNA 9430096L06 gene	845	88
408	AAB29651	Homo sapiens	Human membrane-associated protein HUMAP-8.	502	100
409	gi15074997	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	98	32
409	AAG73357	Homo sapiens	Human gene 12-encoded secreted protein HBXAM53, SEQ ID NO:128.	57	35
409	AAG73405	Homo sapiens	Human gene 12-encoded secreted protein HBXAM53, SEQ ID NO:176.	57	35
410	gi1669689	Homo sapiens	H.sapiens TAFII105 mRNA, partial.	3902	98
410	AAW31494	Homo sapiens	Human hTAFII105 protein.	3902	98
410	AAV57279	Homo sapiens	Transcription factor subunit TAFII105 polypeptide.	3902	98
411	AAG71672	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1353.	1202	94
411	AAG72062	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1743.	1068	66
411	AAG71847	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1528.	1051	67
412	AAV16630	Homo sapiens	Human Putative Adrenomedullin Receptor (PAR).	1592	99
412	gi292419	Homo sapiens	Human homologue of the canine orphan receptor (RDC1) mRNA, 5' end.	1580	98
412	gi899	Canis familiaris	RDC1 receptor (AA 1-362)	1503	92
413	AAV95002	Homo sapiens	Human secreted protein vc34_1, SEQ ID NO:44.	985	71
413	gi14550480	Homo sapiens	clone MGC:16377 IMAGE:3936171, mRNA, complete cds.	917	97
413	gi7020918	Homo sapiens	cDNA FLJ20668 fis, clone KAIA585.	179	37
414	AAB56877	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1455.	1004	98
414	gi13991373	Hymenolepis diminuta	NADH dehydrogenase subunit 4L	62	38
414	gi14487711	Hepatitis C virus	polyprotein	62	50
415	gi179165	Homo sapiens	Human Na,K-ATPase subunit alpha 2 (ATP1A2) gene, complete cds.	5238	99
415	gi203029	Rattus norvegicus	(Na <sup>+</sup> and K <sup>+</sup> ) ATPase, alpha+ catalytic subunit precursor	5205	98
415	gi212406	Gallus gallus	Na,K-ATPase alpha-2-subunit	4977	93
416	AAB90649	Homo sapiens	Human secreted protein, SEQ ID NO: 192.	563	92
416	AAB90565	Homo sapiens	Human secreted protein, SEQ ID NO: 103.	472	100
416	AAB90651	Homo sapiens	Human secreted protein, SEQ ID NO: 194.	203	97
417	gi6599290	Homo sapiens	mRNA; cDNA DKFZp586C1021 (from clone DKFZp586C1021); partial cds.	81	25
417	gi7190652	Chlamydia muridarum	phosphoenolpyruvate-protein phosphotransferase	89	21
417	gi14700035	Aspergillus nidulans	nuclear transport factor 2	76	37
418	gi13249295	Homo sapiens	anion exchanger AE4 mRNA, complete cds.	4951	100

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
418	gi13517508	Homo sapiens	sodium bicarbonate cotransporter (SLC4A9) mRNA, partial cds.	4493	95
418	gi11611537	Oryctolagus cuniculus	anion exchanger 4a	4231	85
419	gi2564913	Homo sapiens	clk2 kinase (CLK2), propin1, cotel1, glucocerebrosidase (GBA), and metaxin genes, complete cds; metaxin pseudogene and glucocerebrosidase pseudogene; and thrombospondin3 (THBS3) gene, partial cds.	1109	82
419	gi1326108	Homo sapiens	Human metaxin (MTX) gene, complete cds.	1109	82
419	gi12804907	Homo sapiens	Similar to metaxin 1, clone MGC:2518 IMAGE:3546178, mRNA, complete cds.	1100	99
420	gi2564913	Homo sapiens	clk2 kinase (CLK2), propin1, cotel1, glucocerebrosidase (GBA), and metaxin genes, complete cds; metaxin pseudogene and glucocerebrosidase pseudogene; and thrombospondin3 (THBS3) gene, partial cds.	1665	100
420	gi1326108	Homo sapiens	Human metaxin (MTX) gene, complete cds.	1665	100
420	gi807670	Mus musculus	metaxin	1519	91
421	gi6094684	Homo sapiens	PAC clone RP1-278D1 from X, complete sequence.	580	30
421	gi7023516	Homo sapiens	cDNA FLJ11078 fis, clone PLACE1005102, weakly similar to RING CANAL PROTEIN.	547	30
421	AAB93480	Homo sapiens	Human protein sequence SEQ ID NO:12768.	547	30
422	gi14715068	Homo sapiens	Similar to RIKEN cDNA 2600001A11 gene, clone MGC:9907 IMAGE:3870073, mRNA, complete cds.	2062	100
422	gi3342906	Homo sapiens	2-amino-3-ketobutyrate-CoA ligase mRNA, nuclear gene encoding mitochondrial protein, complete cds.	853	89
422	gi4093159	Mus musculus	2-amino-3-ketobutyrate-coenzyme A ligase	834	87
423	AAB24058	Homo sapiens	Human PRO290 protein sequence SEQ ID NO:7.	1972	100
423	AAY66639	Homo sapiens	Membrane-bound protein PRO290.	1972	100
423	AAB65162	Homo sapiens	Human PRO290 (UNQ253) protein sequence SEQ ID NO:33.	1972	100
424	gi167835	Dictyostelium discoideum	myosin heavy chain	152	24
424	gi14042847	Homo sapiens	cDNA FLJ14957 fis, clone PLACE4000009, weakly similar to MYOSIN HEAVY CHAIN, NONMUSCLE TYPE B.	135	26
424	AAB95546	Homo sapiens	Human protein sequence SEQ ID NO:18167.	135	26
425	AAB43587	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1032.	427	100
425	AAG00658	Homo sapiens	Human secreted protein, SEQ ID NO:	360	97

136  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			4739.		
425	AAG00657	Homo sapiens	Human secreted protein, SEQ ID NO: 4738.	243	72
426	gi13325388	Homo sapiens	Similar to RIKEN cDNA 1110007C09 gene, clone MGC:11115 IMAGE:3833318, mRNA, complete cds.	535	99
426	AAB93133	Homo sapiens	Human protein sequence SEQ ID NO:12027.	77	30
427	gi7023138	Homo sapiens	cDNA FLJ10847 fis, clone NT2RP4001379.	731	49
427	AAB93249	Homo sapiens	Human protein sequence SEQ ID NO:12263.	731	49
427	AAB18977	Homo sapiens	Amino acid sequence of a human transmembrane protein.	616	89
428	AAB18977	Homo sapiens	Amino acid sequence of a human transmembrane protein.	1008	100
428	gi7023138	Homo sapiens	cDNA FLJ10847 fis, clone NT2RP4001379.	765	43
428	AAB93249	Homo sapiens	Human protein sequence SEQ ID NO:12263.	765	43
429	AAG03349	Homo sapiens	Human secreted protein, SEQ ID NO: 7430.	59	28
429	gi12620543	Bradyrhizobium japonicum	ID263	63	30
429	AAY20368	Homo sapiens	Human microtubule associated protein 2 mutant fragment 64.	53	40
430	gi7209839	Homo sapiens	mRNA for casein kinase I epsilon, complete cds.	1564	99
430	gi13676318	Homo sapiens	casein kinase 1, epsilon, clone MGC:10398 IMAGE:3937782, mRNA, complete cds.	1564	99
430	gi852057	Homo sapiens	casein kinase I epsilon mRNA, complete cds.	1564	99
431	gi2642187	Rattus norvegicus	endo-alpha-D-mannosidase	1973	87
431	gi10434559	Homo sapiens	cDNA FLJ12838 fis, clone NT2RP2003230, moderately similar to Rattus norvegicus endo-alpha-D-mannosidase (Enman) mRNA.	1559	99
431	AAB95204	Homo sapiens	Human protein sequence SEQ ID NO:17303.	1559	99
432	gi12044469	Homo sapiens	mRNA; cDNA DKFZp761H1710 (from clone DKFZp761H1710); complete cds.	141	37
432	gi15079305	Mus musculus	RIKEN cDNA 9130020G10 gene	126	37
432	gi6599277	Homo sapiens	mRNA; cDNA DKFZp434E1818 (from clone DKFZp434E1818); partial cds.	114	41
433	gi12803977	Homo sapiens	clone MGC:4175 IMAGE:3634983, mRNA, complete cds.	611	100
433	AAB34781	Homo sapiens	Human secreted protein sequence encoded by gene 9 SEQ ID NO:69.	58	39
433	AAW39938	Homo sapiens	Peptide effecting G-protein-coupled receptor activity.	57	37
434	gi2150013	Homo sapiens	transmembrane protein mRNA, complete cds.	1159	100

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
434	gi12803197	Homo sapiens	claudin 5 (transmembrane protein deleted in velocardiofacial syndrome), clone MGC:8543 IMAGE:2822745, mRNA, complete cds.	1159	100
434	AAY91533	Homo sapiens	Human secreted protein sequence encoded by gene 83 SEQ ID NO:206.	1159	100
435	gi15082442	Homo sapiens	clone MGC:20235 IMAGE:4562851, mRNA, complete cds.	1368	100
435	gi7023829	Homo sapiens	cDNA FLJ11273 fis, clone PLACE1009338.	503	42
435	AAB93645	Homo sapiens	Human protein sequence SEQ ID NO:13146.	503	42
436	gi11640570	Homo sapiens	MSTP031 mRNA, complete cds.	777	100
436	AAY91516	Homo sapiens	Human secreted protein sequence encoded by gene 66 SEQ ID NO:189.	70	44
436	AAY91657	Homo sapiens	Human secreted protein sequence encoded by gene 66 SEQ ID NO:330.	70	44
437	AAG73464	Homo sapiens	Human gene 7-encoded secreted protein fragment, SEQ ID NO:239.	2267	98
437	AAG73462	Homo sapiens	Human gene 7-encoded secreted protein fragment, SEQ ID NO:237.	1898	99
437	AAG73463	Homo sapiens	Human gene 7-encoded secreted protein fragment, SEQ ID NO:238.	1881	98
438	gi9886738	Homo sapiens	JP3 mRNA for junctophilin type3, complete cds.	3916	99
438	gi9927307	Mus musculus	junctophilin type 3	3549	90
438	gi9886757	Homo sapiens	JP3 gene for junctophilin type3, exon 5 and partial cds.	3172	100
439	AAB08894	Homo sapiens	Human secreted protein sequence encoded by gene 4 SEQ ID NO:51.	240	64
439	gi7414441	porcine endogenous retrovirus	envelope protein	147	28
439	gi348952	Rat leukemia virus	envelope protein	145	26
440	gi13623369	Homo sapiens	clone IMAGE:3957135, mRNA, partial cds.	2617	100
440	AAB43484	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:929.	761	100
440	gi14247685	Staphylococcus aureus subsp. aureus Mu50	nicotinate phosphoribosyltransferase homolog	370	40
441	gi13623369	Homo sapiens	clone IMAGE:3957135, mRNA, partial cds.	2077	94
441	AAB43484	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:929.	761	100
441	gi14247685	Staphylococcus aureus subsp. aureus Mu50	nicotinate phosphoribosyltransferase homolog	370	40
442	gi13623369	Homo sapiens	clone IMAGE:3957135, mRNA, partial cds.	2517	97
442	AAB43484	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:929.	761	100
442	gi14247685	Staphylococcus	nicotinate phosphoribosyltransferase	370	40

138  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		s aureus subsp. aureus Mu50	homolog		
443	gi13182757	Homo sapiens	HTPAP mRNA, complete cds.	639	65
443	AAB70690	Homo sapiens	Human hDPP protein sequence SEQ ID NO:7.	639	65
443	gi14020949	Arabidopsis thaliana	phosphatidic acid phosphatase	460	39
444	gi10436254	Homo sapiens	cDNA FLJ13948 fis, clone Y79AA1001023.	529	41
444	AAB94837	Homo sapiens	Human protein sequence SEQ ID NO:16006.	529	41
444	gi7022187	Homo sapiens	cDNA FLJ10261 fis, clone HEMBB1000975.	521	42
445	gi1403547	Saccharomyces cerevisiae	P2558 protein	162	26
445	gi2621070	Methanothermobacter thermautotrophicus	ribosomal protein S18 (E.coli S13)	79	33
445	gi4097361	Human parainfluenza virus 1	nucleocapsid protein	59	30
446	gi15157363	Agrobacterium tumefaciens	AGR_C_4025p	259	32
446	gi15075368	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	251	31
446	gi15024663	Clostridium acetobutylicum	Uncharacterized protein, YfiH family	198	28
447	gi12584947	Homo sapiens	ovary-specific acidic protein mRNA, complete cds.	1195	100
447	gi632549	Petromyzon marinus	NF-180	152	30
447	gi4678807	Homo sapiens	Human gene from PAC 179D3, chromosome X, isoform of mitochondrial apoptosis inducing factor, AIF, AF100928.	140	32
448	AAX23994_aal	Homo sapiens	Human CAR receptor DNA.	1495	99
448	gi458542	Homo sapiens	H.sapiens mRNA for orphan nuclear hormone receptor.	1494	99
448	AAR41346	Homo sapiens	Human CAR receptor polypeptide.	1494	99
449	gi14625447	Rattus norvegicus	MT-protocadherin	2566	83
449	AAB12154	Homo sapiens	Hydrophobic domain protein isolated from WERI-RB cells.	895	100
449	gi13537202	Homo sapiens	PC-LKC mRNA for protocadherin LKC, complete cds.	445	31
450	gi10880797	Mus musculus	Syne-1A	124	27
450	gi5262574	Homo sapiens	mRNA; cDNA DKFZp434G173 (from clone DKFZp434G173); complete cds.	108	26
450	gi10880799	Mus musculus	Syne-1B	124	27
451	gi11967375	Rattus norvegicus	Dvl-binding protein Idax	1062	100

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
451	gi11967377	Homo sapiens	Dvl-binding protein IDAX mRNA, complete cds.	1062	100
451	gi7023269	Homo sapiens	cDNA FLJ10920 fis, clone OVARC1000384.	348	48
452	gi4929538	Rattus norvegicus	Olig-1 bHLH protein	1088	87
452	gi11602814	Mus musculus	Olig1 bHLH protein	1070	86
452	gi7385152	Mus musculus	oligodendrocyte-specific bHLH transcription factor Olig1	1070	86
453	gi3851514	Phytophthora infestans	cyst germination specific acidic repeat protein precursor	874	31
453	gi454154	Homo sapiens	intestinal mucin (MUC2) mRNA, complete cds.	746	26
453	gi296881	Clostridium thermocellum	S-layer protein	678	34
454	gi4929577	Homo sapiens	CGI-54 protein mRNA, complete cds.	1552	100
454	AAY13942	Homo sapiens	Human transmembrane protein, HP01737.	1552	100
454	AAB36611	Homo sapiens	Human FLEXHT-33 protein sequence SEQ ID NO:33.	1546	99
455	gi295671	Saccharomyces cerevisiae	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III	108	21
455	gi2425111	Dictyostelium discoideum	ZipA	107	20
455	gi1279563	Medicago sativa	nuM1	104	21
456	AAB58236	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 574.	286	88
456	gi2065288	Doryctobracon crawfordi	cytochrome b	61	30
456	gi1653554	Synechocystis sp. PCC 6803	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	48	45
457	gi3273731	Homo sapiens	MHC class I region.	603	95
457	gi312407	Homo sapiens	Human HLA-F gene for human leukocyte antigen F.	603	95
457	gi14349362	Homo sapiens	Similar to major histocompatibility complex, class I, F, clone MGC:15399 IMAGE:4039990, mRNA, complete cds.	599	95
458	AAG71945	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1626.	1106	96
458	AAG71532	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1213.	1104	96
458	AAG71525	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1206.	641	53
459	gi11612079	Homo sapiens	DC-specific transmembrane protein mRNA, complete cds.	2448	100
459	AAE02638	Homo sapiens	Human dendritic cell specific transmembrane protein (DC-STAMP).	2448	100
459	AAB87357	Homo sapiens	Human gene 16 encoded secreted protein HMADJ14, SEQ ID NO:98.	1798	99
460	gi3006230	Homo sapiens	PAC clone RP4-604G5 from 7q22-q31.1, complete sequence.	85	35
460	gi47373	Streptococcus pneumoniae	7 kDa protein	59	42



140  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
460	gi5880698	Nephroselmis olivacea	translational initiation factor 1	57	30
461	AAG73470	Homo sapiens	Human gene 14-encoded secreted protein fragment, SEQ ID NO:245.	699	100
461	gi10436625	Homo sapiens	cDNA FLJ14220 fis, clone NT2RP3003828.	489	53
461	AAB95779	Homo sapiens	Human protein sequence SEQ ID NO:18726.	489	53
462	gi7021367	Drosophila melanogaster	c11.1	522	27
462	gi12724134	Lactococcus lactis subsp. lactis	HYPOTHETICAL PROTEIN	84	33
463	gi7322066	Drosophila sp.	Hls	367	28
463	gi3309579	Rattus norvegicus	A-kinase anchor protein121; AKAP121	155	27
463	gi2072307	Mus musculus	AKAP121	154	27
464	AAB47106	Homo sapiens	Second splice variant of MAPP.	4193	99
464	AAB47105	Homo sapiens	First splice variant of MAPP.	3311	100
464	gi14550175	Mus musculus	ADAM33	2684	72
465	gi14091952	Rattus norvegicus	KIDINS220	324	27
465	gi11321435	Rattus norvegicus	ankyrin repeat-rich membrane-spanning protein	320	27
465	gi6599237	Homo sapiens	mRNA; cDNA DKFZp434F0621 (from clone DKFZp434F0621).	220	27
466	gi9864747	Leishmania major	L165.9	225	35
466	gi3021392	Homo sapiens	H.sapiens mRNA for nuclear protein SDK3, partial.	118	34
466	gi5734402	Homo sapiens	mRNA for GANP protein.	96	27
467	gi12002028	Homo sapiens	brain my040 protein mRNA, complete cds.	482	100
467	AAB56147	Homo sapiens	Human secreted protein sequence encoded by gene 71 SEQ ID NO:241.	74	36
467	AAB56272	Homo sapiens	Human secreted protein sequence encoded by gene 71 SEQ ID NO:366.	74	36
468	AAAY94938	Homo sapiens	Human secreted protein clone ye78_1 protein sequence SEQ ID NO:82.	2290	97
468	gi13603412	Homo sapiens	B29 mRNA, complete cds.	187	30
468	AAAY17227	Homo sapiens	Human secreted protein (clone ya1-1).	203	26
469	AAAY27721	Homo sapiens	Human secreted protein encoded by gene No. 29.	1118	88
469	AAB87068	Homo sapiens	Human secreted protein TANGO 365, SEQ ID NO:46.	621	99
469	AAB87146	Homo sapiens	Human secreted protein TANGO 365 ASV variant, SEQ ID NO:161.	617	98
470	gi10438739	Homo sapiens	cDNA: FLJ22376 fis, clone HRC07327.	1931	99
470	AAE03639	Homo sapiens	Human extracellular matrix and cell adhesion molecule-3 (XMAD-3).	1934	99
470	gi4033606	Adiantum capillus-veneris	Extensin	200	33
471	gi1769467	Homo sapiens	Human p126 (ST5) mRNA, complete cds.	1504	70

141

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
471	gi1769472	Homo sapiens	Human p82 (ST5) mRNA, alternatively spliced, complete cds.	1504	70
471	gi257387	human, revertant clone F2, mRNA Partial, 2687 nt]. [Homo sapiens	HTS1=HeLa tumor suppressor gene	1504	70
472	gi9944535	Amsacta moorei entomopoxvirus	AMV012	69	29
472	gi559500	Caenorhabditis elegans	ND2 protein (AA 1 - 282)	81	35
472	gi15042251	Chilo iridescent virus	150R	62	36
473	gi559500	Caenorhabditis elegans	ND2 protein (AA 1 - 282)	91	26
473	gi9944535	Amsacta moorei entomopoxvirus	AMV012	69	29
473	gi9944642	Amsacta moorei entomopoxvirus	AMV119	73	29
474	gi5739566	Homo sapiens	BAC clone CTA-332P12 from 7q22-q31.1, complete sequence.	907	100
474	gi32474	Homo sapiens	H.sapiens h-Sp1 mRNA.	907	100
474	gi632790	human, keratinocyte line HaCaT, mRNA, 2106 nt]. [Homo sapiens	pantophysin	907	100
475	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene, clone MGC:19636 IMAGE:2822323, mRNA, complete cds.	937	100
475	AAB36613	Homo sapiens	Human FLEXHT-35 protein sequence SEQ ID NO:35.	937	100
475	gi7022832	Homo sapiens	cDNA FLJ10661 fis, clone NT2RP2006106.	240	90
476	gi5052674	Drosophila melanogaster	BcDNA.LD29892	162	38
476	AAB21007	Homo sapiens	Human nucleic acid-binding protein, NuABP-11.	167	39
476	gi9295345	Homo sapiens	HSKM-B (HSKM-B) mRNA, complete cds.	173	31
477	AAG71509	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1190.	1510	96
477	AAG71669	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1350.	1198	77
477	AAG71820	Homo sapiens	Human olfactory receptor polypeptide,	1181	75

142  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SEQ ID NO: 1501.		
478	AAAY73483	Homo sapiens	Human secreted protein clone yl18_1 protein sequence SEQ ID NO:188.	582	47
478	AAW85723	Homo sapiens	Novel protein (Clone AX56 28).	246	34
478	AAG03191	Homo sapiens	Human secreted protein, SEQ ID NO: 7272.	112	30
479	gi15079907	Homo sapiens	Similar to secretory carrier membrane protein 4, clone MGC:19661 IMAGE:3161979, mRNA, complete cds.	1182	94
479	gi9837305	Rattus norvegicus	secretory carrier membrane protein 4	1012	79
479	gi7021484	Mus musculus	secretory carrier membrane protein 4	1006	77
480	gi1345560	Oryza sativa	nitrate reductase apoenzyme (AA 394-471) (130 is 2nd base in codon)	72	44
481	gi13517508	Homo sapiens	sodium bicarbonate cotransporter (SLC4A9) mRNA, partial cds.	5138	100
481	gi14582760	Homo sapiens	anion exchanger AE4 mRNA, complete cds.	4603	96
481	gi11611537	Oryctolagus cuniculus	anion exchanger 4a	4080	86
482	gi2570933	Rattus norvegicus	vanilloid receptor subtype 1	986	44
482	gi7544146	Rattus norvegicus	vanilloid receptor type 1 like protein 1	979	45
482	gi11055318	Rattus norvegicus	vanilloid receptor-related osmotically activated channel	951	43
483	gi14669436	Homo sapiens	alkaline phytoceramidase (APHC) mRNA, complete cds.	110	54
483	AAB18986	Homo sapiens	Amino acid sequence of a human transmembrane protein.	110	54
483	gi14488266	Arabidopsis thaliana	Acyl-CoA independent ceramide synthase	91	33
484	gi12053091	Homo sapiens	mRNA; cDNA DKFZp434F1719 (from clone DKFZp434F1719); complete cds.	615	97
484	AAE01546	Homo sapiens	Human gene 1 encoded secreted protein HMVCQ82, SEQ ID NO:96.	76	39
484	gi1574439	Haemophilus influenzae Rd	leucine responsive regulatory protein (lrp)	77	36
485	AAAY99347	Homo sapiens	Human PRO1113 (UNQ556) amino acid sequence SEQ ID NO:24.	2250	99
485	AAB71863	Homo sapiens	Human h15571 GPCR.	1834	48
485	gi7407148	Homo sapiens	protocadherin Flamingo 2 (FMI2) mRNA, complete cds.	306	27
486	AAW94654	Homo sapiens	G-protein coupled receptor HM74A protein.	887	52
486	gi219867	Homo sapiens	Human mRNA for HM74.	882	52
486	AAAY90637	Homo sapiens	Human G protein-coupled receptor HM74.	882	52
487	gi3337385	Homo sapiens	Chromosome 16 BAC clone CIT987SK-A-761H5, complete sequence.	1158	83
487	gi2342743	Homo sapiens	Human Chromosome 16 BAC clone CIT987SK-A-589H1, complete sequence.	709	59
487	gi4959568	Homo sapiens	nuclear pore complex interacting protein NPIP (NPIP) mRNA, complete cds.	705	58
488	gi7021167	Homo sapiens	cDNA FLJ20839 fis, clone ADKA02346.	551	98

143  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
488	gi9309293	Homo sapiens	hasc-1 mRNA for asc-type amino acid transporter 1, complete cds.	551	98
488	gi7415938	Mus musculus	asc1	460	83
489	gi14248997	Homo sapiens	lung seven transmembrane receptor 1 (LUSTR1) mRNA, complete cds.	2239	97
489	gi10439034	Homo sapiens	cDNA: FLJ22591 fis, clone HSI03124.	1515	98
489	gi14248999	Mus musculus	lung seven transmembrane receptor 2	813	49
490	AAAY87079	Homo sapiens	Human secreted protein sequence SEQ ID NO:118.	927	82
490	gi3851540	Homo sapiens	brain mitochondrial carrier protein-1 (BMCP1) mRNA, nuclear gene encoding mitochondrial protein, complete cds.	927	82
490	gi11094335	Homo sapiens	mitochondrial uncoupling protein 5 long form mRNA, complete cds; nuclear gene for mitochondrial product.	927	82
491	AAG71803	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1484.	1616	100
491	AAG71807	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1488.	1165	69
491	AAG71805	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1486.	1099	83
492	gi10440458	Homo sapiens	mRNA for FLJ00065 protein, partial cds.	992	100
492	gi938175	Gallus gallus	alpha1 (XIV) collagen	102	32
492	gi211358	Gallus gallus	alpha-1 collagen type IX	63	45
493	gi9963845	Homo sapiens	HT017 mRNA, complete cds.	558	38
493	AAW09405	Homo sapiens	Pineal gland specific gene-1 protein.	558	38
493	AAB69185	Homo sapiens	Human hSLR-iso protein SEQ ID NO:7.	558	38
494	gi6179740	Homo sapiens	paraneoplastic neuronal antigen MA3 (MA3) mRNA, complete cds.	421	51
494	gi12053257	Homo sapiens	mRNA; cDNA DKFZp434K225 (from clone DKFZp434K225); complete cds.	421	51
494	AAB12529	Homo sapiens	Human Ma5 protein SEQ ID NO:13.	421	51
495	gi13384467	Caenorhabditis elegans	contains similarity to CDP-alcohol phosphotransferases	391	35
495	gi3661595	Arabidopsis thaliana	aminoalcoholphosphotransferase	411	32
495	gi530088	Glycine max	aminoalcoholphosphotransferase	410	31
496	gi9963853	Homo sapiens	HT018 mRNA, complete cds.	1368	100
496	AAG71359	Homo sapiens	Human gene 10-encoded secreted protein fragment, SEQ ID NO:210.	50	50
496	AAAY20863	Homo sapiens	Human presenilin 1 mutant protein fragment 9.	61	36
497	gi13241761	Homo sapiens	transmembrane protein induced by tumor necrosis factor alpha (TMPIT) mRNA, complete cds.	1286	70
497	AAB12123	Homo sapiens	Hydrophobic domain protein from clone HP10608 isolated from Saos-2 cells.	1286	70
497	AAB38371	Homo sapiens	Human secreted protein encoded by gene 51 clone HLDQC46.	331	67
498	AAAY86234	Homo sapiens	Human secreted protein HNTNC20, SEQ ID NO:149.	126	32
498	AAB24074	Homo sapiens	Human PRO1153 protein sequence SEQ ID NO:49.	113	54
498	AAAY66735	Homo sapiens	Membrane-bound protein PRO1153.	113	54

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
499	AAB93704	Homo sapiens	Human protein sequence SEQ ID NO:13287.	3677	99
499	gi2792496	Rattus norvegicus	tulip 2	1339	70
499	gi2792494	Rattus norvegicus	tulip 1	1159	48
500	gi10438718	Homo sapiens	cDNA: FLJ22362 fis, clone HRC06544.	1224	100
500	gi310897	Thermobifida fusca	beta-1,4-endoglucanase precursor	138	36
500	AA Y59066	Homo sapiens	Human tie receptor FNIII repeat fragment 2.	99	26
501	gi4519607	Homo sapiens	Nurr1 gene, complete cds.	1342	100
501	gi4760535	Homo sapiens	gene for T-cell nuclear receptor NOT (Nurr1), complete cds.	1342	100
501	gi14424530	Homo sapiens	nuclear receptor subfamily 4, group A, member 2, clone MGC:14354 IMAGE:4298967, mRNA, complete cds.	1342	100
502	gi7288872	Rattus norvegicus	taste receptor rT2R6	398	32
502	gi7262617	Homo sapiens	candidate taste receptor T2R9 gene, complete cds.	397	33
502	AAB87739	Homo sapiens	Human T2R09 amino acid sequence SEQ ID NO:17.	397	33
503	gi7022610	Homo sapiens	cDNA FLJ10521 fis, clone NT2RP2000841.	3005	98
503	AAB92909	Homo sapiens	Human protein sequence SEQ ID NO:11539.	3005	98
503	gi13111772	Homo sapiens	clone MGC:2899 IMAGE:3010245, mRNA, complete cds.	649	99
504	AAB51244	Homo sapiens	Human haemopoietin receptor protein NR10.3 SEQ ID NO:17.	3066	99
504	AAB51242	Homo sapiens	Human haemopoietin receptor protein NR10.1 SEQ ID NO:2.	3018	100
504	AAB51243	Homo sapiens	Human haemopoietin receptor protein NR10.2 SEQ ID NO:4.	885	100
505	AAG71668	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1349.	1547	97
505	AAG71507	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1188.	1399	90
505	AAG71676	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1357.	1126	70
506	gi10438252	Homo sapiens	cDNA: FLJ22009 fis, clone HEP07114.	2022	99
506	gi12654279	Homo sapiens	clone IMAGE:3451160, mRNA, partial cds.	1975	100
506	gi4102877	Mus musculus	Shc binding protein	1915	70
507	gi12248917	Homo sapiens	mRNA for spinesin, complete cds.	1404	100
507	AAB11699	Homo sapiens	Human serine protease BSSP2 (hBSSP2), SEQ ID NO:10.	1404	100
507	AAB08950	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:107.	1207	100
508	gi7715916	Mus musculus	SorCSb splice variant of the VPS10 domain receptor SorCS	4966	96
508	gi6692583	Mus musculus	VPS10 domain receptor protein SORCS	4961	96
508	gi12007720	Mus musculus	VPS10 domain receptor protein SorCS2	2613	49

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
509	gi10566471	Mus musculus	Gliacolin	1284	94
509	gi14278927	Mus musculus	gliacolin	1284	94
509	gi3747097	Homo sapiens	Clq-related factor mRNA, complete cds.	974	71
510	gi7332063	Caenorhabditis elegans	contains similarity to Strongylocentrotus purpuratus Spec3 protein (SP:P16537)	147	41
510	gi12247892	Sterkiella histriomuscorum	SPEC3-like protein	85	36
510	gi483822	Gallus gallus	vitellogenin gene-binding protein, alpha/alpha isoform	73	47
511	AAB25755	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:144.	648	100
511	AAB25754	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:143.	301	100
511	AAB25697	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:86.	278	100
512	gi13810306	Homo sapiens	mRNA for transmembrane protein 7 (TMEM7 gene).	1271	100
512	gi11065721	Homo sapiens	mRNA for 28kD interferon responsive protein (IFRG28 gene).	420	45
512	AAB84453	Homo sapiens	Amino acid sequence of a human interferon-alpha induced protein.	420	45
513	AAG72504	Homo sapiens	Human OR-like polypeptide query sequence, SEQ ID NO: 2185.	1615	99
513	AAG71709	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1390.	1611	99
513	AAG72127	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1808.	829	99
514	AAB83079	Homo sapiens	Human CASB6411 protein.	1806	100
514	AAB08764	Homo sapiens	A human leukocyte and blood related protein (LBAP).	1424	100
514	gi10435645	Homo sapiens	cDNA FLJ13593 fis, clone PLACE1009493.	1124	100
515	AAB74716	Homo sapiens	Human membrane associated protein MEMAP-22.	1094	99
515	gi6093235	Homo sapiens	mRNA; cDNA DKFZp566N034 (from clone DKFZp566N034); partial cds.	424	94
515	gi15157430	Agrobacterium tumefaciens	AGR_C_4131p	131	25
516	gi13447610	Homo sapiens	VTS20631 mRNA, g-protein coupled receptor family, partial cds.	3804	99
516	gi10441732	Homo sapiens	leucine-rich repeat-containing G protein-coupled receptor 6 (LGR6) mRNA, partial cds.	3782	100
516	gi3366802	Homo sapiens	orphan G protein-coupled receptor HG38 mRNA, complete cds.	1805	52
517	AAB24465	Homo sapiens	Human secreted protein sequence encoded by gene 29 SEQ ID NO:90.	447	98
517	gi1749851	Human immunodeficiency virus type 1	tat protein	60	36
517	gi2245481	Human immunodeficiency virus type 1	Tat protein	59	33

146  
Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		ncy virus type 1			
518	gi5802879	Homo sapiens	AIM-1 protein mRNA, complete cds.	458	44
518	gi15028433	Mus musculus	B/AIM-1-like protein	453	45
518	gi4680229	Homo sapiens	DNb-5 mRNA, partial cds.	498	41
519	gi5525078	Rattus norvegicus	seven transmembrane receptor	788	31
519	AAV57288	Homo sapiens	Human GPCR protein (HGPRP) sequence (clone ID 3036563).	752	29
519	AAV40440	Homo sapiens	Human brain-derived G-protein coupled receptor protein.	746	29
520	AAV27577	Homo sapiens	Human secreted protein encoded by gene No. 11.	598	100
520	gi1617316	Homo sapiens	H.sapiens mRNA for tenascin-R.	97	26
520	gi4379056	Homo sapiens	H.sapiens mRNA for tenascin-R (restrictin).	97	26
521	gi10434488	Homo sapiens	cDNA FLJ12791 fis, clone NT2RP2001991, highly similar to SODIUM- AND CHLORIDE-DEPENDENT TRANSPORTER NTT73.	1523	100
521	AAB94304	Homo sapiens	Human protein sequence SEQ ID NO:14767.	1523	100
521	gi11907841	Homo sapiens	orphan neurotransmitter transporter v7-3 mRNA, complete cds.	1353	92
522	gi10437307	Homo sapiens	cDNA: FLJ21240 fis, clone COL01132.	677	38
522	AAV94906	Homo sapiens	Human secreted protein clone rb649_3 protein sequence SEQ ID NO:18.	644	37
522	AAB74730	Homo sapiens	Human membrane associated protein MEMAP-36.	644	37
523	AAB43665	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1110.	1254	100
523	AAV19759	Homo sapiens	SEQ ID NO 477 from WO9922243.	966	100
523	gi12804249	Homo sapiens	Similar to gene rich cluster, C9 gene, clone MGC:2519 IMAGE:3546861, mRNA, complete cds.	411	46
524	AAB03625	Homo sapiens	Human G-protein coupled receptor fb41a.	1925	94
524	AAB70143	Homo sapiens	Human G protein-coupled receptor protein.	1925	94
524	AAW79258	Homo sapiens	Human G protein coupled receptor 15E.	1877	93
525	gi7023154	Homo sapiens	cDNA FLJ10856 fis, clone NT2RP4001547.	943	53
525	AAV28810	Homo sapiens	nn296_2 secreted protein.	943	53
525	AAB93258	Homo sapiens	Human protein sequence SEQ ID NO:12282.	943	53
526	gi11878036	Sus scrofa	somatostatin receptor 1	198	25
526	gi12056166	Yaba-like disease virus	7L protein	196	26
526	gi13876663	lumpy skin disease virus	G-protein-coupled chemokine receptor-like protein	197	25
527	gi3880799	Caenorhabditis elegans	Y39A1B.2	441	24
527	gi1707052	Caenorhabditis elegans	similar to drosophila and mouse patched proteins	368	23
527	gi1255388	Caenorhabditis	similar to drosophila membrane protein	191	23

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		<i>elegans</i>	PATCHED (SP: P18502)		
528	AAB34321	<i>Homo sapiens</i>	Human secreted protein sequence encoded by gene 23 SEQ ID NO:82.	74	38
528	AAB51693	<i>Homo sapiens</i>	Human secreted protein related amino acid sequence SEQ ID NO:133.	51	55
528	AAB87388	<i>Homo sapiens</i>	Human gene 47 encoded secreted protein HFXDK20, SEQ ID NO:129.	68	44
529	AAAY94297	<i>Homo sapiens</i>	Human coenzyme A-utilising enzyme CoAEN-5.	1581	69
529	AAAY66699	<i>Homo sapiens</i>	Membrane-bound protein PRO1108.	1581	69
529	AAB65222	<i>Homo sapiens</i>	Human PRO1108 (UNQ551) protein sequence SEQ ID NO:248.	1581	69
530	AAAY29332	<i>Homo sapiens</i>	Human secreted protein clone pe584_2 protein sequence.	1282	99
530	AAB58289	<i>Homo sapiens</i>	Lung cancer associated polypeptide sequence SEQ ID 627.	1282	99
530	AAB75246	<i>Homo sapiens</i>	Human secreted protein sequence encoded by gene 7 SEQ ID NO:65.	1282	99
531	AAB08538	<i>Homo sapiens</i>	A human G-protein coupled receptor designated 14273.	787	100
531	AAAY44662	<i>Homo sapiens</i>	Human 14273 G-protein coupled receptor (GPCR).	765	98
531	AAAY44815	<i>Homo sapiens</i>	Human 14273 G-protein coupled receptor (GPCR) version 2.	761	97
532	AAG71706	<i>Homo sapiens</i>	Human olfactory receptor polypeptide, SEQ ID NO: 1387.	1579	99
532	AAG71705	<i>Homo sapiens</i>	Human olfactory receptor polypeptide, SEQ ID NO: 1386.	1180	74
532	AAG71679	<i>Homo sapiens</i>	Human olfactory receptor polypeptide, SEQ ID NO: 1360.	1089	68
533	gi557822	<i>Saccharomyces cerevisiae</i>	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	362	27
533	gi1304387	<i>Saccharomyces cerevisiae</i> var. diastaticus	glucoamylase	362	27
533	gi7332056	<i>Caenorhabditis elegans</i>	contains similarity to Pfam family PF00078 (Reverse transcriptase (RNA-dependent)), score=79.6, E=6.3e-20, E=1	345	27
534	AAU00437	<i>Homo sapiens</i>	Human dendritic cell membrane protein FIRE.	1841	91
534	AAAY91625	<i>Homo sapiens</i>	Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	1840	90
534	AAAY59300	<i>Homo sapiens</i>	Human EGPCR polypeptide.	1121	58
535	gi10438710	<i>Homo sapiens</i>	cDNA: FLJ22357 fis, clone HRC06404.	4572	100
535	gi14336678	<i>Homo sapiens</i>	16p13.3 sequence section 1 of 8.	4547	99
535	AAB61148	<i>Homo sapiens</i>	Human NOV17 protein.	1955	67
536	gi10438710	<i>Homo sapiens</i>	cDNA: FLJ22357 fis, clone HRC06404.	4379	100
536	gi14336678	<i>Homo sapiens</i>	16p13.3 sequence section 1 of 8.	4354	99
536	AAB61148	<i>Homo sapiens</i>	Human NOV17 protein.	1955	67
537	gi10439790	<i>Homo sapiens</i>	cDNA: FLJ23186 fis, clone LNG11945.	753	99
537	gi310100	<i>Rattus norvegicus</i>	developmentally regulated protein	86	30
537	gi5824457	<i>Caenorhabditis</i>	contains similarity to Pfam domain:	78	30



Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		elegans	PF00615 (Regulator of G protein signaling domain), Score=200.4, E-value=9.1e-57, N=1		
538	AAG71899	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1580.	1603	100
538	gi5869925	Mus musculus	olfactory receptor	1322	82
538	AAG71954	Homo sapiens	Human olfactory receptor polypeptide, SEQ ID NO: 1635.	883	54
539	gi466604	Escherichia coli	No definition line found	90	25
539	gi52952	Mus musculus	delta-aminolevulinate dehydratase (AA 1 - 330)	82	35
539	gi4262032	Bos taurus	D5 dopamine receptor	59	64
540	gi12803977	Homo sapiens	clone MGC:4175 IMAGE:3634983, mRNA, complete cds.	611	100
540	AAB34781	Homo sapiens	Human secreted protein sequence encoded by gene 9 SEQ ID NO:69.	58	39
540	AAW39938	Homo sapiens	Peptide effecting G-protein-coupled receptor activity.	57	37
541	AAV73442	Homo sapiens	Human secreted protein clone ya66_1 protein sequence SEQ ID NO:106.	596	95
541	AAB63255	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:617.	95	40
541	gi13182890	Macaca mulatta	collagen type III alpha 1	79	46
542	gi9929914	Homo sapiens	MUC3B gene for intestinal mucin, partial cds.	4024	99
542	gi9929918	Homo sapiens	MUC3B mRNA for intestinal mucin, partial cds.	4024	99
542	gi11990203	Homo sapiens	partial MUC3B gene for MUC3B mucin, exons 1-11.	3985	98
543	gi14043332	Homo sapiens	Similar to ring finger protein 23, clone MGC:2475 IMAGE:3051389, mRNA, complete cds.	925	40
543	gi10716078	Mus musculus	testis-abundant finger protein	919	40
543	gi12407417	Mus musculus	tripartite motif protein TRIM11	671	35
544	gi57131	Rattus norvegicus	ribosomal protein S26	260	68
544	gi12803549	Homo sapiens	ribosomal protein S26, clone MGC:1963 IMAGE:3143099, mRNA, complete cds.	260	68
544	gi456351	Homo sapiens	H.sapiens RPS26 mRNA.	260	68
545	gi10438861	Homo sapiens	cDNA: FLJ22461 fis, clone HRC10107.	1258	42
545	gi15079400	Homo sapiens	clone MGC:16796 IMAGE:3855477, mRNA, complete cds.	1258	42
545	gi6683905	Drosophila melanogaster	Dispatched	412	37
546	AAV72910	Homo sapiens	Human IGS3 G-protein coupled receptor (GPCR) protein.	589	58
546	AAB67654	Homo sapiens	Amino acid sequence of a human G-protein coupled receptor (Ant).	589	58
546	AAF55661 aal	Homo sapiens	Nucleotide sequence of a human G-protein coupled receptor (Ant).	589	58
547	gi6740013	Homo sapiens	clone cDSC1 Down syndrome cell adhesion molecule (DSCAM) mRNA,	6373	60

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			complete cds.		
547	AAW42086	Homo sapiens	Human Down syndrome-cell adhesion molecule DS-CAM1.	6347	62
547	gi11066998	Mus musculus	Down syndrome cell adhesion molecule	6344	60
548	gi12656633	Homo sapiens	transmembrane gamma-carboxyglutamic acid protein 3 TMG3 mRNA, complete cds.	1192	100
548	gi2338290	Homo sapiens	proline-rich Gla protein 1 (PRGP1) mRNA, complete cds.	283	49
548	gi506601	Rattus norvegicus	factor X	206	49
549	gi12698682	Homo sapiens	testis-expressed transmembrane-4 protein (TETM4) mRNA, complete cds.	588	95
549	gi11559214	Homo sapiens	mRNA for MS4A5, complete cds.	588	95
549	gi13649401	Homo sapiens	MS4A5 protein mRNA, complete cds.	588	95
550	gi12054393	Homo sapiens	6M1-10*01 gene for olfactory receptor, cell line BM28.7.	1853	100
550	gi12054395	Homo sapiens	6M1-10*01 gene for olfactory receptor, cell line BM19.7.	1853	100
550	gi12054397	Homo sapiens	6M1-10*01 gene for olfactory receptor, cell line LG2.	1853	100
551	gi11275360	Homo sapiens	SLC4A10 mRNA for NCBE, complete cds.	5677	99
551	gi11182364	Mus musculus	NCBE	5542	96
551	gi7385123	Mus musculus	sodium bicarbonate cotransporter isoform 3 kNBC-3	4364	76
552	AAE04178	Homo sapiens	Human gene 3 encoded secreted protein fragment, SEQ ID NO:169.	1111	98
552	AAE04127	Homo sapiens	Human gene 3 encoded secreted protein HSDJL42, SEQ ID NO:114.	1078	98
552	AAE04102	Homo sapiens	Human gene 3 encoded secreted protein HSDJL42, SEQ ID NO:88.	1068	98

150  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
277	AAY55787	Homo sapiens	INCY- Human zinc RING (ZIRI) protein.	1859	95
277	AAW81821	Homo sapiens	INCY- Human ZIRI protein.	1859	95
277	gi3387925	Homo sapiens	RING zinc finger protein RZF	1859	95
278	AAY55787	Homo sapiens	INCY- Human zinc RING (ZIRI) protein.	1703	88
278	AAW81821	Homo sapiens	INCY- Human ZIRI protein.	1703	88
278	gi3387925	Homo sapiens	RING zinc finger protein RZF	1703	88
279	AAY55787	Homo sapiens	INCY- Human zinc RING (ZIRI) protein.	1769	92
279	AAW81821	Homo sapiens	INCY- Human ZIRI protein.	1769	92
279	gi3387925	Homo sapiens	RING zinc finger protein RZF	1769	92
280	AAB24463	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 27 SEQ ID NO:88.	1346	96
280	AAU27674	Homo sapiens	ZYMO Human protein AFP669232.	1334	95
280	AAB34813	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 41 SEQ ID NO:101.	701	93
281	ABB89737	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2113.	614	87
281	AAG89173	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 293.	614	87
281	AAM25811	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1326.	614	87
282	AAW61622	Homo sapiens	HUMA- Clone HTPBA27 of TM4SF superfamily.	841	93
282	gi2997747	Homo sapiens	tetraspan TM4SF; Tspan-4	841	93
282	gi2586350	Homo sapiens	tetraspan	841	93
283	gi15080477	Homo sapiens	Similar to RIKEN cDNA 2310010G13 gene	2034	97
283	gi17512422	Mus musculus	Similar to RIKEN cDNA 2310010G13 gene	1577	76
283	gi17427162	Ralstonia solanacearum	TRANSPORT TRANSMEMBRANE PROTEIN	315	28
284	ABB05645	Homo sapiens	BODE- Human thyroglobulin 38 protein SEQ ID NO:2.	1858	100
284	ABB05646	Homo sapiens	BODE- Human thyroglobulin 38 protein N-terminal peptide SEQ ID NO:7.	88	100
284	gi21322795	Corynebacterium glutamicum ATCC 13032	ABC-type transporter, permease components	78	22
285	gi18157547	Mus musculus	pecanex-like 3	1791	93
285	gi15076843	Homo	pecanex-like protein 1	871	34

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens			
285	AAM42412	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 145.	743	100
286	gi17390957	Mus musculus	Similar to RIKEN cDNA 2010001E11 gene	184	26
286	gi2650264	Archaeoglobus fulgidus	oxalate/formate antiporter (oxlT-2)	95	22
286	gi19712705	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	Multidrug resistance protein 2	94	18
287	AAW27484	Homo sapiens	IMUT- Human MCP.	1991	96
287	gi180137	Homo sapiens	membrane cofactor protein	1991	96
287	AAR93939	Homo sapiens	AUST- CD46 wild-type.	1986	96
288	AAE01687	Homo sapiens	HUMA- Human gene 16 encoded secreted protein HDPMM88, SEQ ID NO:99.	1019	100
288	AAO14187	Homo sapiens	INCY- Human transporter and ion channel TRICH-4.	560	58
288	gi20988041	Homo sapiens	Similar to ATPase, Class I, type 8B, member 2	560	58
289	AAG81436	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:390.	392	100
289	AAG74872	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5636.	392	100
289	AAB08863	Homo sapiens	INCY- Amino acid sequence of a human secretory protein.	392	100
290	gi1226246	Homo sapiens	mono-ADP-ribosyltransferase	1880	94
290	gi2677616	Mus musculus	NAD(P)(+)-arginine ADP-ribosyltransferase	1142	60
290	gi20067374	Mus musculus	ART3 mono(ADP-ribosyl)transferase	1071	58
291	AAB70690	Homo sapiens	SREN- Human hDPP protein sequence SEQ ID NO:7.	598	100
291	AAG89279	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 399.	598	100
291	gi13182757	Homo sapiens	HTPAP	598	100
292	AAU83599	Homo sapiens	GETH Human PRO protein, Seq ID No 16.	760	100
292	AAB88418	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0181.	725	100
292	ABK09980_aal	Homo sapiens	JAKO/ Human prostate stem cell antigen (PSCA) cDNA sequence.	101	32
293	gi12718841	Mus musculus	Skullin	279	38
293	gi4191356	Mus musculus	claudin-6	277	38
293	gi13543081	Mus musculus	claudin 6	277	38

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent Identity
294	ABB50276	Homo sapiens	USSH HLA-DR alpha chain ovarian tumour marker protein, SEQ ID NO:41.	1214	92
294	AAB58160	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 498.	1214	92
294	gi15929084	Homo sapiens	major histocompatibility complex, class II, DR alpha	1214	92
295	AAE15283	Homo sapiens	INCY- Human RNA metabolism protein-46 (RMEP-46).	2777	99
295	gi16768810	Drosophila melanogaster	LD05247p	1133	46
295	gi16185327	Drosophila melanogaster	LD38433p	906	40
296	gi12620132	Homo sapiens	renal sodium/sulfate cotransporter	3100	100
296	gi469555	Rattus norvegicus	Na/Sulfate cotransporter	2627	82
296	gi310183	Rattus norvegicus	sodium dependent sulfate transporter	2627	82
297	AAAY44245	Homo sapiens	INCY- Human cell signalling protein-8.	1522	89
297	AAE06590	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10785.	1327	80
297	AAM93721	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3671.	1205	99
298	AAE13277	Homo sapiens	INCY- Human transporters and ion channels (TRICH)-4.	3306	92
298	AAD06381_aal	Homo sapiens	ACT1- Human ATP binding cassette, ABCB9 transporter cDNA.	2338	99
298	AAE02437	Homo sapiens	ACT1- Human ATP binding cassette, ABCB9 transporter protein.	2338	99
299	gi20072551	Mus musculus	RIKEN cDNA 4930511J11 gene	342	40
299	gi17974542	Homo sapiens	voltage-dependent calcium channel gamma-8 subunit	118	25
299	gi13357180	Homo sapiens	calcium channel gamma subunit 8	117	25
300	gi20258606	Homo sapiens	sideroflexin 5	1178	100
300	gi3874886	Caenorhabditis elegans	C41C4.2	592	46
300	gi13543138	Mus musculus	RIKEN cDNA 2810002O05 gene	401	38
301	AAE07054	Homo sapiens	HUMA- Human gene 4 encoded secreted protein HSYAB05, SEQ ID NO:71.	612	29
301	AAE07077	Homo sapiens	HUMA- Human gene 4 encoded secreted protein HSYAB05, SEQ ID NO:94.	143	23
301	gi9964007	Homo sapiens	MAB21L2 protein	105	33
302	ABB89405	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1781.	1337	98
302	gi15030135	Mus musculus	RIKEN cDNA 1110020A09 gene	769	60
302	gi16767870	Drosophila	GH02466p	284	36

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		melanogaster			
303	AAE13349	Homo sapiens	SENO- Human TSTP protein, 165-015D.	1652	100
303	AAE13348	Homo sapiens	SENO- Human TSTP protein, 165-015C.	589	40
303	AAE13350	Homo sapiens	SENO- Human TSTP protein, 165-015E.	314	31
304	ABB89737	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2113.	489	100
304	AAG89173	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 293.	489	100
304	AAM25811	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1326.	489	100
305	gi16648454	Drosophila melanogaster	SD01285p	290	30
305	AAY87336	Homo sapiens	INCY- Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	222	28
305	gi4877582	Homo sapiens	lipoma HMGIC fusion partner	222	28
306	AAE14439	Homo sapiens	INCY- Human drug metabolising enzyme (DME)-2.	1123	98
306	ABB84932	Homo sapiens	GETH Human PRO3579 protein sequence SEQ ID NO:232.	1123	98
306	AAB87576	Homo sapiens	GETH Human PRO3579.	1123	98
307	gi18857903	Homo sapiens	TCBA1	867	100
307	AAG78000	Homo sapiens	BIOW- Human actin 14.	663	100
307	ABB89045	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1421.	644	98
308	gi4580997	Mus musculus	cAMP inducible 2 protein	2377	87
308	gi18676548	Homo sapiens	FLJ00171 protein	1877	100
308	gi20073163	Mus musculus	Similar to solute carrier family 37 (glycerol-3-phosphate transporter), member 1	1572	60
309	AAG71797	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1478.	755	100
309	AAG66336	Homo sapiens	CURA- Human NOV 16 protein sequence.	755	100
309	AAU24615	Homo sapiens	SENO- Human olfactory receptor AOLFR108.	755	100
311	AAS01280_aa1	Homo sapiens	JANC Human alpha nicotinic acetylcholine receptor cDNA sequence.	2370	100
311	AAD27812_aa1	Homo sapiens	GLAX Human nicotinic acetylcholine receptor gene, sbg471005nAChR.	2370	100
311	AAE17317	Homo sapiens	GLAX Human nicotinic acetylcholine receptor protein, sbg471005nAChR.	2370	100
312	gi21518639	Homo sapiens	TSLC1-like 2	1991	97
312	gi19068139	Mus musculus	membrane glycoprotein	1970	96

154  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
312	AAM78418	Homo sapiens	HYSE- Human protein SEQ ID NO 1080.	1905	97
313	AAG67512	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	3994	100
313	AAH78215_aal	Homo sapiens	SMIK Nucleotide sequence of a human secreted polypeptide.	1659	57
313	AAG67523	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	1659	57
314	ABB90749	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 230.	2691	100
314	ABB90723	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 179.	2691	100
314	gi15987487	Homo sapiens	tumor endothelial marker 3 precursor	2691	100
315	ABB90749	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 230.	2600	97
315	ABB90723	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 179.	2600	97
315	gi15987487	Homo sapiens	tumor endothelial marker 3 precursor	2600	97
316	AAG66705	Homo sapiens	CURA- Human GPCR3 polypeptide.	1494	100
316	AAG71567	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1248.	1414	100
316	gi18480740	Mus musculus	olfactory receptor MOR267-14	1017	67
317	AAU83597	Homo sapiens	GETH Human PRO protein, Seq ID No 12.	690	31
317	ABB10293	Homo sapiens	HUMA- Human cDNA SEQ ID NO: 601.	651	100
317	ABB10483	Homo sapiens	HUMA- Human cDNA SEQ ID NO: 791.	642	99
318	gi10944274	Homo sapiens	bA346K17.2 (A novel protein similar to the cell division control protein 91 (CDC91, YLR459W or L9122.2) from Yeast)	2235	100
318	gi20988986	Homo sapiens	CDC91 cell division cycle 91-like 1 (S. cerevisiae)	2235	100
318	AAB88430	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0205.	2226	99
319	AAV19506	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	1120	100
319	gi 17540010 ref NP_503066.1	Caenorhabditis elegans	F26D10.11.p	83	28
319	gi 14149748 ref NP_068365.1	Mus musculus	claudin 15	72	20
320	gi784990	Homo sapiens	5-HT5A serotonin receptor	1645	100
320	gi20379144	Homo sapiens	5-hydroxytryptamine receptor 5A	1645	100
320	AAR45848	Homo sapiens	INRM Human 5HT5a serotonin receptor.	1611	98
321	AAS07947_aal	Homo sapiens	AREN- Human cDNA encoding G-protein coupled receptor, hRUP20.	1734	100

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent Identity
321	AAD13260_aal	Homo sapiens	MILL- Human 39406 cDNA.	1734	100
321	AAM50774	Homo sapiens	INGE- Human G protein coupled receptor IGPCr20.	1734	100
322	AAV25806	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 23.	1663	98
322	gi19528215	Drosophila melanogaster	AT30101p	1012	38
322	AAM93717	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3663.	1011	100
323	AAB12119	Homo sapiens	PROT- Hydrophobic domain protein from clone HP02869 isolated from KB cells.	448	100
323	gi4827164	Gluconacetobacter xylinus	similar to melibiose carrier protein of E.coli	89	26
323	gi595475	Homo sapiens	hFcRn	84	31
324	AAV25736	Homo sapiens	HUMA- Human secreted protein encoded from gene 26.	343	100
325	AAB44336	Homo sapiens	HUMA- Human secreted protein encoded by gene 2 clone HROAM11.	169	100
325	gi12045265 ref NP_073076.1	Mycoplasma genitalium	ATP synthase F0, subunit B (atpF)	65	44
325	gi18447301 gb AAL68225.1	Drosophila melanogaster	LD26265p	65	31
326	gi14278927	Mus musculus	gliacolin	1291	94
326	gi10566471	Mus musculus	Gliacolin	1291	94
326	gi3747097	Homo sapiens	C1q-related factor	976	70
327	gi13506225	Mus musculus	ST7 protein form1 splice variant a	2996	99
327	gi19353275	Mus musculus	Similar to suppression of tumorigenicity 7	2940	98
327	gi9230665	Homo sapiens	FAM4A1 splice variant a	2857	95
328	gi9230665	Homo sapiens	FAM4A1 splice variant a	2709	94
328	gi13506227	Mus musculus	ST7 protein form1 splice variant b	2702	94
328	gi13506225	Mus musculus	ST7 protein form1 splice variant a	2668	90
329	gi9230667	Homo sapiens	FAM4A1 splice variant b	2859	99
329	gi13506225	Mus musculus	ST7 protein form1 splice variant a	2848	96
329	gi19353275	Mus musculus	Similar to suppression of tumorigenicity 7	2792	95
330	AAU19222	Homo sapiens	PHAA Human G protein-coupled receptor nGPCR-2343.	467	100
330	AAV25491	Homo	BGHM cDNA for Epstein Barr virus	317	38



156  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	aal	sapiens	induced gene 2 (EBI-2).		
330	AAAY90630	Homo sapiens	AREN- Human G protein-coupled receptor EBI2.	317	38
331	AAB94231	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14604.	3584	99
331	AAB95784	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18737.	3570	100
331	gi10880791	Homo sapiens	PP791 protein	3329	99
332	AAAY23325	Homo sapiens	GETH A33 related antigen JAM.	105	27
332	gi3462455	Mus musculus	junctional adhesion molecule	105	27
332	gi8650528	Rattus norvegicus	junctional adhesion molecule JAM	98	26
333	AAG93279	Homo sapiens	NISC- Human protein HP03145.	1977	99
333	gi14250676	Homo sapiens	Similar to RIKEN cDNA 2310002F18 gene	1977	99
333	AAAY27589	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 23.	1578	100
334	gi953239	Homo sapiens	tetraspan membrane protein	996	91
334	gi12655071	Homo sapiens	transmembrane 4 superfamily member 4	996	91
334	gi11493837	Rattus norvegicus	tetraspan protein LRTM4	911	81
335	AAB94238	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14621.	3039	99
335	AAB87342	Homo sapiens	HUMA- Human gene 1 encoded secreted protein HETHR73, SEQ ID NO:83.	3033	99
335	AAU23815	Homo sapiens	UROG- Human prostate-related gene 103P2D6 encoded protein.	3016	99
336	gi14336694	Homo sapiens	M83	4100	99
336	gi18204292	Homo sapiens	transmembrane protein 8 (five membrane-spanning domains)	4096	99
336	gi10716072	Homo sapiens	M83 protein	4089	99
337	AAD02700_aal	Homo sapiens	REGC Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.	2056	100
337	AAE15438	Homo sapiens	INCY- Human drug metabolising enzyme (DME)-5.	2056	100
337	AAAY72640	Homo sapiens	REGC Human glycosyl sulfotransferase-4beta (GST-4beta).	2056	100
338	AAB82971	Homo sapiens	MILL- G protein coupled receptor 43238.	1631	99
338	gi18480770	Mus musculus	olfactory receptor MOR271-1	1373	83
338	gi18479336	Mus musculus	olfactory receptor MOR270-1	1367	83
339	AAB82971	Homo sapiens	MILL- G protein coupled receptor 43238.	1562	99
339	gi18479336	Mus musculus	olfactory receptor MOR270-1	1338	85

157

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
339	gi18480770	Mus musculus	olfactory receptor MOR271-1	1336	84
340	gi7960136	Homo sapiens	neurexin 3 isoform	4557	100
340	gi1145791	Rattus norvegicus	neurexin 3	4505	98
340	gi7960135	Homo sapiens	neurexin 3 isoform	4419	97
341	ABB07253	Homo sapiens	LEXI- Human novel GPCR (NGPCR) protein.	3943	99
341	AAM69607	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29913.	1770	82
341	AAM57201	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29306.	1770	82
342	AAG72315	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1996.	1140	76
342	AAE18020	Homo sapiens	CURA- Human G-protein coupled receptor-7 (GPCR-7) protein.	915	96
342	AAU24629	Homo sapiens	SENO- Human olfactory receptor AOLFR123.	859	89
343	AAB95124	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17122.	1552	81
343	gi854065	Human herpesvirus 6	U88	802	46
343	AAM40934	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5865.	435	36
344	AAG71823	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1504.	1627	100
344	AAU24669	Homo sapiens	SENO- Human olfactory receptor AOLFR167.	1627	100
344	AAE11910	Homo sapiens	CURA- Human G-protein coupled receptor 15a (GPCR15a) protein.	1627	100
345	AAU00437	Homo sapiens	COUN- Human dendritic cell membrane protein FIRE.	2867	88
345	AAV91625	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	1966	97
345	gi16930385	Mus musculus	seven-span membrane protein FIRE	1838	55
346	AAU00437	Homo sapiens	COUN- Human dendritic cell membrane protein FIRE.	2341	87
346	AAV91625	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	1966	97
346	gi16930385	Mus musculus	seven-span membrane protein FIRE	1535	59
347	ABB94047	Homo sapiens	HUMA- Human secreted protein SEQ ID NO: 90.	84	31
347	ABB94023	Homo sapiens	HUMA- Human secreted protein SEQ ID NO: 66.	84	31
347	gi21288752 gb EAA01045.1	Anopheles gambiae str. PEST	ebiP7790	537	34
348	AAW75000	Homo sapiens	HUMA- Human secreted protein encoded by gene 146 clone HSNK17.	349	100
348	ABB03792	Homo	HUMA- Human musculoskeletal system	70	28

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	related polypeptide SEQ ID NO 1739.		
348	gi 17542842 ref NP_500310.1	Caenorhabditis elegans	W08E12.8.p	69	39
349	gi19684136	Homo sapiens	Similar to RIKEN cDNA 4933413N12 gene	178	26
349	gi841378	Saccharomyces cerevisiae	Gpi2p	90	30
349	gi295139	Staphylococcus lugdunensis	ORFB	79	31
350	AAB88406	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0162.	1421	99
350	ABB50346	Homo sapiens	HUMA- Human secreted protein encoded by gene 46 SEQ ID NO:294.	476	95
350	AAW88579	Homo sapiens	HUMA- Secreted protein encoded by gene 46 clone HCFMV39.	476	95
351	gi292793	Homo sapiens	T-cell receptor beta	636	98
351	AAM76093	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36399.	594	93
351	AAM63281	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35386.	594	93
352	AAV10839	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	225	95
353	AAV16784	Homo sapiens	GEMY Human secreted protein (clone co1000 1).	488	100
353	gi1850866	Macropus robustus	ATPase subunit 8	69	31
353	gi2935032	Rhodococcus opacus	ClcR	68	42
354	gi 21293186 gb EAA05331.1	Anopheles gambiae str. PEST	agCP9246	71	26
355	AAA40083_aa1	Homo sapiens	KAZU- Human brain-specific transmembrane glycoprotein encoding cDNA.	1553	51
355	AAB12448	Homo sapiens	CHUG- Human hh00149 protein SEQ ID NO:4.	1553	51
355	AAB09968	Homo sapiens	KAZU- Human brain-specific transmembrane glycoprotein.	1553	51
356	AAB50953	Homo sapiens	GETH Human PRO534 protein.	1760	95
356	AAB73689	Homo sapiens	INCY- Human oxidoreductase protein ORP-22.	1760	95
356	AAB44303	Homo sapiens	GETH Human PRO534 (UNQ335) protein sequence SEQ ID NO:410.	1760	95
357	gi12276180	Homo sapiens	metalloprotease-disintegrin meltrin beta	5255	99
357	AAE19181	Homo sapiens	INCY- Human protease, PRPS-18 protein.	4967	99
357	gi12802370	Homo sapiens	disintegrin and metalloproteinase ADAM19	4967	99
358	gi18056675	Homo	FREB	1969	98

Table 2B

SEQ ID	Hlt ID	Species	Description	S score	Percent identity
		sapiens			
358	gi21245136	Homo sapiens	FCRLa1	1940	99
358	AAE03451	Homo sapiens	HUMA- Human gene 25 encoded secreted protein HRGBL78, SEQ ID NO: 134.	1888	98
359	gi18056675	Homo sapiens	FREB	1986	99
359	AAE03451	Homo sapiens	HUMA- Human gene 25 encoded secreted protein HRGBL78, SEQ ID NO: 134.	1905	99
359	AAB34744	Homo sapiens	ALPH- Human secreted protein encoded by DNA clone vq24 1.	1905	99
360	AAW74807	Homo sapiens	HUMA- Human secreted protein encoded by gene 79 clone HSKNE46.	270	100
360	AAO02082	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15974.	69	41
360	AAB34697	Homo sapiens	ALPH- Human secreted protein encoded by DNA clone vq6 1.	66	45
361	gi17861418	Drosophila melanogaster	GH03649p	226	35
361	gi6959684	Mus musculus	glycolipid transfer protein	95	24
361	gi16741551	Mus musculus	Similar to glycolipid transfer protein	95	24
362	AAE06578	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10769.	2337	100
362	gi13623231	Homo sapiens	Similar to RIKEN cDNA 1200013A08 gene	2337	100
362	AAB92464	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10520.	2272	98
363	AAU12211	Homo sapiens	GETH Human PRO1886 polypeptide sequence.	1639	99
363	gi17542564 ref NP_501434.1	Caenorhabditis elegans	T26A8.2.p	189	21
363	gi21298000 gb EAA10145.1	Anopheles gambiae str. PEST	agCP15426	127	18
364	ABB05715	Homo sapiens	GEHU- Human transmembrane protein clone tes3 17i21.	1237	100
364	AAU27674	Homo sapiens	ZYMO Human protein AFP669232.	649	48
364	AAB24463	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 27 SEQ ID NO:88.	648	48
365	gi14582572	Homo sapiens	orphan transporter SLC19A3	2549	100
365	gi12483888	Homo sapiens	solute carrier 19A3	2549	100
365	gi12483890	Mus musculus	solute carrier 19A3	1713	68
366	AAM41254	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6185.	632	90
366	ABB11854	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2224.	632	90
366	ABB89257	Homo	HUMA- Human polypeptide SEQ ID NO	631	99

160  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	1633.		
367	AAB94138	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14406.	2598	100
367	gi15866720	Homo sapiens	fukutin-related protein	2598	100
367	gi17945162	Drosophila melanogaster	RE09574p	354	23
368	AAE14448	Homo sapiens	INCY- Human drug metabolising enzyme (DME)-11.	2002	99
368	AAB85780	Homo sapiens	INCY- Human drug metabolizing enzyme (ID No. 7256116CD1).	1797	98
368	gi4519535	Homo sapiens	Leukotriene B4 omega-hydroxylase	1222	64
369	gi18157547	Mus musculus	pecanex-like 3	1809	95
369	gi15076843	Homo sapiens	pecanex-like protein 1	872	34
369	AAM42412	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 145.	743	100
370	AAB61219	Homo sapiens	MILL- Human TANGO 292 protein.	1201	100
370	gi14603178	Homo sapiens	transmembrane gamma-carboxylglutamic acid protein 4	1201	100
370	gi12656635	Homo sapiens	transmembrane gamma-carboxylglutamic acid protein 4 TMG4	1201	100
371	AAM40584	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5515.	2045	95
371	ABB10286	Homo sapiens	HUMA- Human cDNA SEQ ID NO: 594.	2045	95
371	ABB10269	Homo sapiens	HUMA- Human cDNA SEQ ID NO: 577.	2045	95
372	gi1510143	Homo sapiens	similar to C.elegans protein encoded in cosmid T20D3 (Z68220).	1624	55
372	ABB89128	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1504.	1359	98
372	AAV53635	Homo sapiens	CHIR A bone marrow secreted protein designated BMS53.	1148	51
373	AAB93444	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12686.	1006	87
373	ABB89562	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1938.	998	86
373	gi15209353	Caenorhabditis elegans	Y39B6A.1	138	45
374	AAM06271	Homo sapiens	HYSE- Human foetal protein, SEQ ID NO: 2.	426	98
374	gi190203	Homo sapiens	potassium channel	76	32
374	gi10176968	Arabidopsis thaliana	receptor-like protein kinase	76	31
375	gi5542014	Homo sapiens	dyskerin	2616	91
375	AAV33675	Homo sapiens	DEKR- Human DKC1 protein.	2549	90
375	gi3135028	Homo	dyskerin	2549	90

161  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens			
376	gi5542014	Homo sapiens	dyskerin	2492	94
376	AAY33675	Homo sapiens	DEKR- Human DKC1 protein.	2425	92
376	gi3135028	Homo sapiens	dyskerin	2425	92
377	gi1763011	Homo sapiens	lysophospholipase homolog	1444	90
377	gi13623261	Homo sapiens	lysophospholipase-like	1444	90
377	gi14594904	Homo sapiens	monoglyceride lipase	1390	90
378	gi1763011	Homo sapiens	lysophospholipase homolog	1589	92
378	gi13623261	Homo sapiens	lysophospholipase-like	1589	92
378	gi14594904	Homo sapiens	monoglyceride lipase	1535	92
379	ABB90165	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2541.	571	93
379	AAY94946	Homo sapiens	GEMY Human secreted protein clone cd205 2 protein sequence SEQ ID NO:98.	571	93
379	AAY53051	Homo sapiens	GEMY Human secreted protein clone dd119 4 protein sequence SEQ ID NO:108.	318	59
380	AAM93503	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3213.	1082	92
380	AAY77122	Homo sapiens	INCY- Human neurotransmission-associated protein (NTAP) 414692.	1082	92
380	gi6523817	Homo sapiens	SIR protein	1082	92
381	AAE07124	Homo sapiens	HUMA- Human gene 16 encoded secreted protein fragment, SEQ ID NO:141.	931	91
381	AAE07099	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO:116.	931	91
381	gi6980032	Mus musculus	ARL-6 interacting protein-1	907	88
382	gi21430284	Drosophila melanogaster	LD38689p	1292	40
382	AAM80289	Homo sapiens	HYSE- Human protein SEQ ID NO 3935.	191	30
382	AAM79305	Homo sapiens	HYSE- Human protein SEQ ID NO 1967.	191	30
383	AAG73684	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4448.	1863	98
383	AAY48312	Homo sapiens	META- Human prostate cancer-associated protein 9.	1509	100
383	gi17389322	Homo sapiens	Similar to NICE-5 protein	1419	74
384	AAB93185	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12134.	2492	100
384	AAM93581	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3373.	1971	96
384	AAE10328	Homo	INCY- Human transporter and ion channel-5	1873	100

162  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	(TRICH-5) protein.		
385	ABB89951	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2327.	2862	99
385	AAB58984	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 692.	759	94
385	ABB04610	Homo sapiens	BODA- Human quinoprotein dehydrogenase 33 protein SEQ ID NO:2.	244	27
386	ABB89951	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2327.	2791	98
386	AAB58984	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 692.	688	89
386	ABB04610	Homo sapiens	BODA- Human quinoprotein dehydrogenase 33 protein SEQ ID NO:2.	251	28
387	AAM93354	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2907.	531	100
387	AAM00917	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 393.	495	99
387	gi18308220	Xenopus laevis	transmembrane protein quicken	333	77
388	AAU12232	Homo sapiens	GETH Human PRO4398 polypeptide sequence.	2696	100
388	ABB90111	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2487.	1784	99
388	gi14860862	Homo sapiens	polyamine oxidase isoform-1	932	39
389	AAM00947	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 423.	6659	98
389	AAM00834	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 197.	4723	100
389	AAY99666	Homo sapiens	INCY- Human GTPase associated protein-17.	3647	97
390	AAE17492	Homo sapiens	INCY- Human secretion and trafficking protein-1 (SAT-1).	1705	100
390	gi13529623	Mus musculus	Similar to RIKEN cDNA 4930418P06 gene	1408	81
390	gi 21313292 ref NP_084053.1	Mus musculus	RIKEN cDNA 4930418P06	1401	80
391	AAB36613	Homo sapiens	INCY- Human FLEXHT-35 protein sequence SEQ ID NO:35.	1121	85
391	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene	1121	85
391	AAB93042	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11827.	240	90
392	AAB82940	Homo sapiens	UYNH Human androgen receptor trapped protein 5 (ART5).	299	39
392	AAB56085	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 9 SEQ ID NO:179.	299	39
392	gi18043859	Mus musculus	Similar to RIKEN cDNA 9430098E02 gene	251	42
393	AAM39990	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3135.	1209	70
393	AAM38999	Homo	HYSE- Human polypeptide SEQ ID NO	1209	70

163  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	2144.		
393	AAB18993	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	1209	70
394	gi4220892	Homo sapiens	transcriptional co-activator CRSP34	919	97
394	gi7141322	Homo sapiens	p37 TRAP/SMCC/PC2 subunit	918	97
394	gi16741439	Mus musculus	RIKEN cDNA 1500015J03 gene	918	97
395	gi1825729	Caenorhabditis elegans	C. elegans PTR-2 protein (corresponding sequence C32E8.8)	1024	30
395	gi3880799	Caenorhabditis elegans	Y39A1B.2	940	29
395	gi15718594	Caenorhabditis elegans	C. elegans PTR-10 protein (corresponding sequence F55F8.1)	818	28
396	AAB20342	Homo sapiens	UYMC- Peroxisome proliferator-activated receptor alpha.	2265	94
396	AAR74053	Homo sapiens	LIGA- Human peroxisome proliferator activated receptor.	2265	94
396	gi765240	Homo sapiens	peroxisome proliferator activated receptor alpha; PPAR alpha	2265	94
397	ABB11934	Homo sapiens	HYSE- Human transmembrane protein homologue, SEQ ID NO:2304.	1692	100
397	AAB43983	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1428.	1692	100
397	AAH47123_aa1	Homo sapiens	NIGE- Human B1466 protein encoding cDNA.	1409	100
398	gi19526687	Mus musculus	Na-H exchanger isoform NHE8	2829	96
398	gi5304871	Homo sapiens	dJ963K23.4 (continues in dJ1041C10 (AL162615))	2236	100
398	gi17862784	Drosophila melanogaster	LP02993p	1535	55
399	AAB93258	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12282.	1617	99
399	AAV28810	Homo sapiens	GEMY nn296_2 secreted protein.	1617	99
399	ABB89196	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1572.	1319	99
400	AAG00388	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4469.	316	100
401	AAU21958	Homo sapiens	HUMA- Human cardiovascular system antigen polypeptide SEQ ID No 732.	97	26
401	gi1814196	Caenorhabditis elegans	AO13 ankyrin	87	31
401	gi19110782	Homo sapiens	DNA helicase HEL308	81	25
402	gi21438549	Homo sapiens	humane cDNA	2566	99
402	gi21438547	Rattus norvegicus	Ratten cDNA	2444	93
402	gi21438551	Mus musculus	genomische DNA Exon I der Maus	691	91
403	AAE04759	Homo	INCY- Human vesicle trafficking protein-2	1013	100



164  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	(VETRP-2) protein.		
403	AAB98207	Homo sapiens	SHAN- Human P24 protein-22 SEQ ID NO:2.	1009	99
403	gi16118876	Homo sapiens	vesicular membrane protein P24	1009	99
404	ABB14761	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 3418.	873	95
404	AAU25439	Homo sapiens	INCY- Human mddt protein from clone LG:403872.1:2000MAY19.	524	38
404	AAU75787	Homo sapiens	INCY- Human protein phosphatase 5 (PP5) protein sequence.	444	36
405	AAM93259	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2709.	1257	100
405	gi16877659	Homo sapiens	Similar to RIKEN cDNA 1810054O13 gene	1157	98
405	AAG81420	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:358.	137	40
406	gi12214288	Homo sapiens	dJ402H5.2 (novel protein similar to worm and fly proteins)	1397	50
406	gi3880799	Caenorhabditis elegans	Y39A1B.2	707	25
406	gi1825729	Caenorhabditis elegans	C. elegans PTR-2 protein (corresponding sequence C32E8.8)	602	24
407	gi19338984	Homo sapiens	fat cell-specific low molecular weight protein beta	135	44
407	gi19071802	Homo sapiens	fat cell-specific low molecular weight protein alpha	135	44
407	gi20380358	Mus musculus	RIKEN cDNA 1110025G12 gene	121	31
408	ABB90225	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2601.	952	100
408	AAB12150	Homo sapiens	PROT- Hydrophobic domain protein isolated from HT-1080 cells.	952	100
408	ABB06157	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:249.	944	98
409	gi15074997	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	96	32
409	gi20868002 ref XP_137398.1	Mus musculus	similar to expressed sequence AW049604	75	28
410	AAV57279	Homo sapiens	YEDA Transcription factor subunit TAFII105 polypeptide.	3902	98
410	AAW31494	Homo sapiens	REGC Human hTAFII105 protein.	3902	98
410	gi1669689	Homo sapiens	TBP associated factor	3902	98
411	AAE04639	Homo sapiens	MILL- Human novel transmembrane protein, 32164 protein.	1588	98
411	AAE18658	Homo sapiens	INCY- Human G-protein coupled receptor (GCRC-19).	1548	98
411	AAG71672	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1353.	1202	94
412	ABB11920	Homo sapiens	HYSE- Human adrenomedullin receptor homologue, SEQ ID NO:2290.	1795	95
412	AAV16630	Homo	SMIK Human Putative Adrenomedullin	1789	94

165

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	Receptor (PAR).		
412	gi292419	Homo sapiens	orphan receptor	1774	93
413	AAY95002	Homo sapiens	ALPH- Human secreted protein vc34_1, SEQ ID NO:44.	1027	56
413	ABB12222	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2592.	697	76
413	AAM95374	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4032.	477	65
414	ABB89474	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1850.	1004	98
414	AAB56877	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1455.	1004	98
414	gi18044902	Mus musculus	Similar to RIKEN cDNA 3110005G23 gene	851	65
415	gi179165	Homo sapiens	Na,K-ATPase subunit alpha 2	5238	99
415	gi203029	Rattus norvegicus	(Na <sup>+</sup> and K <sup>+</sup> ) ATPase, alpha+ catalytic subunit precursor	5205	98
415	gi212406	Gallus gallus	Na,K-ATPase alpha-2-subunit	4977	93
416	gi18606367	Mus musculus	RIKEN cDNA 4930570C03 gene	715	92
416	AAB90649	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO: 192.	562	97
416	AAB90565	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO: 103.	472	100
417	gi18512192	Homo sapiens	polycystic kidney and hepatic disease 1	1871	100
417	gi178273	Homo sapiens	alanine:glyoxylate aminotransferase	77	26
417	gi28561	Homo sapiens	L- alanine:glyoxylate aminotransferase	77	26
418	gi13249295	Homo sapiens	anion exchanger AE4	4951	100
418	gi7363254	Homo sapiens	sodium bicarbonate cotransporter 5	4898	98
418	gi13517508	Homo sapiens	sodium bicarbonate cotransporter	4873	95
419	gi2564913	Homo sapiens	metaxin	1108	82
419	gi12804907	Homo sapiens	Similar to metaxin 1	1100	99
419	gi807670	Mus musculus	metaxin	995	89
420	gi2564913	Homo sapiens	metaxin	1665	100
420	gi18606009	Mus musculus	metaxin	1528	91
420	gi12804907	Homo sapiens	Similar to metaxin 1	1470	90
421	gi6094684	Homo sapiens	similar to Kelch proteins; similar to BAA77027 (PID:g4650844)	694	31
421	AAB93480	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12768.	630	29

166  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
421	AAU28187	Homo sapiens	HYSE- Novel human secretory protein, Seq ID No 356.	628	29
422	gi14715068	Homo sapiens	Similar to RIKEN cDNA 2600001A11 gene	2062	100
422	gi4808241	Homo sapiens	dJ466N1.2 (glycine C-acetyltransferase (2-amino-3-ketobutyrate coenzyme A ligase))	853	89
422	gi3342906	Homo sapiens	2-amino-3-ketobutyrate-CoA ligase	853	89
423	AAB65162	Homo sapiens	GETH Human PRO290 (UNQ253) protein sequence SEQ ID NO:33.	1972	100
423	AAV66639	Homo sapiens	GETH Membrane-bound protein PRO290.	1972	100
423	AAB24058	Homo sapiens	GETH Human PRO290 protein sequence SEQ ID NO:7.	1972	100
424	gi167835	Dictyostelium discoideum	myosin heavy chain	142	24
424	gi2983243	Aquifex aeolicus	chromosome assembly protein homolog	140	20
424	AAB95546	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18167.	132	25
425	AAB43587	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1032.	427	100
425	AAM52659	Homo sapiens	BIOW- Human phosphatase 9.	423	98
425	AAG00658	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4739.	360	97
426	gi13325388	Homo sapiens	Similar to RIKEN cDNA 1110007C09 gene	821	88
426	ABB89804	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2180.	814	87
426	AAG73935	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4699.	299	95
427	AAB93249	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12263.	731	49
427	AAB18977	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	615	89
427	AAE01518	Homo sapiens	HUMA- Human gene 2 encoded secreted protein fragment, SEQ ID NO:175.	495	98
428	AAB18977	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	1008	100
428	AAB93249	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12263.	756	43
428	AAV00276	Homo sapiens	HUMA- Human secreted protein encoded by gene 19.	603	100
430	gi7644318	Mesocricetus auratus	casein kinase I epsilon; CKI epsilon	1564	99
430	gi13122442	Rattus norvegicus	casein kinase I epsilon-2	1564	99
430	gi9650968	Rattus norvegicus	casein kinase I epsilon-3	1564	99
431	gi2642187	Rattus norvegicus	endo-alpha-D-mannosidase	1973	87
431	AAB95204	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17303.	1559	99

167  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
431	AAE04255	Homo sapiens	HUMA- Human gene 4 encoded secreted protein fragment, SEQ ID NO:116.	1408	98
432	ABB05662	Homo sapiens	GEHU- Human signal transduction protein clone amy2_10h17.	139	36
432	AAU16313	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1266.	139	36
432	gi21040537	Homo sapiens	Similar to RIKEN cDNA 9130020G10 gene	132	35
433	AAG89209	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 329.	460	97
433	gi1890812	Flexamia graminea	NADH dehydrogenase 1	71	24
433	gi 21295981 gb EAA08126.1	Anopheles gambiae str. PEST	agCP1281	73	28
434	AAAY91533	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 83 SEQ ID NO:206.	1159	100
434	gi2150013	Homo sapiens	transmembrane protein	1159	100
434	gi12803197	Homo sapiens	claudin 5 (transmembrane protein deleted in velocardiofacial syndrome)	1159	100
435	AAE06609	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10800.	498	42
435	ABB89766	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2142.	497	42
435	AAB93645	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13146.	497	42
436	gi11640570	Homo sapiens	MSTP031	777	100
436	ABB50826	Homo sapiens	HUMA- Human secreted protein encoded by gene 77 SEQ ID NO:779.	75	40
436	gi15291231	Drosophila melanogaster	GH13214p	72	25
437	AAG73464	Homo sapiens	HUMA- Human gene 7-encoded secreted protein fragment, SEQ ID NO:239.	2264	98
437	AAG73462	Homo sapiens	HUMA- Human gene 7-encoded secreted protein fragment, SEQ ID NO:237.	1897	100
437	AAG73463	Homo sapiens	HUMA- Human gene 7-encoded secreted protein fragment, SEQ ID NO:238.	1878	98
438	gi9886738	Homo sapiens	junctophilin type3	3916	99
438	gi9927307	Mus musculus	junctophilin type 3	3551	90
438	gi9886757	Homo sapiens	junctophilin type3	3172	100
439	ABB89241	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1617.	739	96
439	gi18762530	Danio rerio	envelope protein	380	47
439	AAB08894	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 4 SEQ ID NO:51.	240	64
440	AAB43484	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:929.	761	100
440	gi10834676	Homo sapiens	PP3856	673	99

168  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
440	gi21428806	Drosophila melanogaster	GH04243p	636	49
441	AAB43484	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:929.	761	100
441	gi21428806	Drosophila melanogaster	GH04243p	636	49
441	gi14247685	Staphylococcus aureus subsp. aureus Mu50	nicotinate phosphoribosyltransferase homolog	544	34
442	AAB43484	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:929.	761	100
442	gi21428806	Drosophila melanogaster	GH04243p	636	49
442	gi10834676	Homo sapiens	PP3856	582	89
443	ABB11177	Homo sapiens	HYSE- Human phosphatidate phosphohydrolase homologue, SEQ ID NO:1547.	952	98
443	AAG89279	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 399.	641	66
443	AAB70690	Homo sapiens	SREN- Human hDPP protein sequence SEQ ID NO:7.	639	65
444	AAM40391	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3536.	672	48
444	AAM42177	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 7108.	567	49
444	ABB90382	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2758.	559	42
445	gi19354040	Mus musculus	Similar to-RIKEN cDNA 1810038N08 gene	853	95
445	gi1403547	Saccharomyces cerevisiae	P2558 protein	175	26
445	AAE15269	Homo sapiens	INCY- Human RNA metabolism protein-32 (RMEP-32).	78	28
446	gi15157363	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_C_4025p	256	31
446	gi15075368	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	243	31
446	gi21324924	Corynebacterium glutamicum ATCC 13032	Uncharacterized ACR	192	28
447	gi20069113	Homo sapiens	corneal endothelium specific protein 1	1201	100
447	gi12584947	Homo sapiens	ovary-specific acidic protein	1195	100

169  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens			
447	gi15214757	Mus musculus	Similar to RIKEN cDNA 4930583H14 gene	558	50
448	AAT92305_aal	Homo sapiens	SALK Constitutively active receptor-alpha encoding cDNA.	1686	94
448	AAG63170	Homo sapiens	TULA- Amino acid sequence of human CAR-a polypeptide.	1686	94
448	AAW93902	Homo sapiens	GEHO Human CAR receptor protein.	1686	94
449	gi18182375	Bos taurus	photoreceptor cadherin	2693	86
449	gi14625447	Rattus norvegicus	MT-protocadherin	2563	83
449	gi18182377	Mus musculus	photoreceptor cadherin	2561	83
450	AAM39421	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2566.	126	27
450	gi18676458	Homo sapiens	FLJ00126 protein	126	27
450	gi17861384	Homo sapiens	nesprin-2 gamma	126	27
451	gi11967375	Rattus norvegicus	Dvl-binding protein Idax	1062	100
451	gi11967377	Homo sapiens	Dvl-binding protein IDAX	1062	100
451	ABB16307	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 4964.	1006	100
452	gi20073201	Homo sapiens	Similar to Olg-1 bHLH protein	1301	100
452	gi4929538	Rattus norvegicus	Olg-1 bHLH protein	1086	87
452	gi7385152	Mus musculus	oligodendrocyte-specific bHLH transcription factor Olig1	1069	86
453	AAM68085	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28391.	6900	99
453	AAM55707	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 27812.	6900	99
453	gi18146660	Homo sapiens	DPCR1	1206	100
454	AAG75611	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6375.	1759	89
454	AAY13942	Homo sapiens	SAGA Human transmembrane protein, HP01737.	1759	89
454	gi15559308	Homo sapiens	Similar to serologically defined breast cancer antigen 84	1759	89
455	gi15430296	Mus musculus	heart alpha-kinase	100	24
455	gi602255	Rattus norvegicus	protein tyrosine phosphatase 2E	99	22
455	gi2425111	Dictyostelium discoideum	ZipA	94	20
456	AAB58236	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 574.	283	88
457	gi5420183	Homo sapiens	dJ377H14.9 (major histocompatibility complex, class I, F (CDA12))	611	96

170  
Table 2B

SEQ ID	Hlt ID	Species	Description	S score	Percent identity
457	AAG64617	Homo sapiens	KIMU/ Human cancer cell specific HLA-F antigen SEQ ID 4.	603	95
457	ABB50296	Homo sapiens	USSH HLA-Cw ovarian tumour marker protein, SEQ ID NO:82.	603	95
458	AAE18015	Homo sapiens	CURA- Human G-protein coupled receptor-3 (GPCR-3) protein.	1116	97
458	AAU24535	Homo sapiens	SENO- Human olfactory receptor AOLFR20.	1116	97
458	AAG71945	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1626.	1106	96
459	AAE02638	Homo sapiens	SCHE Human dendritic cell specific transmembrane protein (DC-STAMP).	2448	100
459	gi11612079	Homo sapiens	DC-specific transmembrane protein	2448	100
459	AAB87357	Homo sapiens	HUMA- Human gene 16 encoded secreted protein HMADJ14, SEQ ID NO:98.	1798	99
460	ABB89120	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1496.	403	87
460	gi17742567	dipeptide	ABC transporter, membrane spanning protein [Agrobacterium tumefaciens str. C58 (U.	71	29
460	gi15159154	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_L_1477p	71	29
461	AAG73470	Homo sapiens	HUMA- Human gene 14-encoded secreted protein fragment, SEQ ID NO:245.	699	100
461	ABB90038	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2414.	486	53
461	AAB95779	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18726.	486	53
462	gi7021367	Drosophila melanogaster	c11.1	511	25
462	gi17862452	Drosophila melanogaster	LD28902p	511	25
462	gi12724134	Lactococcus lactis subsp. lactis	HYPOTHETICAL PROTEIN	81	33
463	AAM42407	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 140.	606	100
463	AAM95921	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4579.	606	100
463	gi7322066	Drosophila sp.	Hls	335	27
464	gi18147612	Homo sapiens	metalloprotease disintegrin	4206	100
464	AAB47106	Homo sapiens	ZYMO Second splice variant of MAPP.	4190	99
464	gi13157560	Homo sapiens	dJ964F7.1 (novel disintegrin and reprolysin metalloproteinase family protein)	4104	100
465	gi14091952	Rattus norvegicus	KIDINS220	294	26

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
465	gi11321435	Rattus norvegicus	ankyrin repeat-rich membrane-spanning protein	292	26
465	AAM39025	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2170.	288	27
466	gi16648368	Drosophila melanogaster	LD35341p	177	49
466	gi19744967	Dictyostelium discoideum	80 kda MCM3-associated protein	153	22
466	gi4995703	Mus musculus	GANP protein	141	25
467	gi12002028	Homo sapiens	brain my040 protein	482	100
467	gi 20453865 gb AAM22167.1 AF482520.1	Utricularia geminiscapa	cytochrome C oxidase subunit I	67	48
467	gi 20453861 gb AAM22165.1 AF482518.1	Utricularia adpressa	cytochrome C oxidase subunit I	67	48
468	AA94938	Homo sapiens	GEMY Human secreted protein clone ye78_1 protein sequence SEQ ID NO:82.	2288	97
468	AAG81379	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:276.	1701	99
468	AAG81387	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:292.	1570	99
469	AA927721	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 29.	1114	98
469	AAB87068	Homo sapiens	MILL- Human secreted protein TANGO 365, SEQ ID NO:46.	621	99
469	AAB87148	Homo sapiens	MILL- Human secreted protein TANGO 365 T20S variant, SEQ ID NO:165.	617	98
470	gi12140288	Homo sapiens	ba12M19.1.3 (novel protein)	2537	100
470	gi12140289	Homo sapiens	ba12M19.1.1 (novel protein)	2203	88
470	AAE03639	Homo sapiens	INCY- Human extracellular matrix and cell adhesion molecule-3 (XMAD-3).	2114	88
471	AAR90766	Homo sapiens	USSH Tumour suppressor protein HTS-1.	1502	70
471	gi257387	Homo sapiens	HTS1	1502	70
471	gi1769472	Homo sapiens	p82	1502	70
472	gi19684136	Homo sapiens	Similar to RIKEN cDNA 4933413N12 gene	645	100
472	gi559500	Caenorhabditis elegans	ND2 protein (AA 1 - 282)	75	35
472	gi6687124	Convolvulus arvensis	NADH dehydrogenase subunit F	72	30
473	gi19684136	Homo sapiens	Similar to RIKEN cDNA 4933413N12 gene	972	100
473	gi2258350	Reclinomonas	SecY-type transporter protein	78	24



172  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		as americana			
473	gi559500	Caenorhabditis elegans	ND2 protein (AA 1 - 282)	76	29
474	gi32474	Homo sapiens	h-Spl	1250	93
474	gi632790	Homo sapiens	pantophysin	1250	93
474	gi16877127	Homo sapiens	Similar to synaptophysin-like protein	1161	92
475	AAB36613	Homo sapiens	INCY- Human FLEXHT-35 protein sequence SEQ ID NO:35.	1304	88
475	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene	1304	88
475	AAB93042	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11827.	240	90
476	gi5052674	Drosophila melanogaster	BcDNA.LD29892	349	24
476	gi16768704	Drosophila melanogaster	HL04910p	329	24
476	gi17945748	Drosophila melanogaster	RE32936p	277	22
477	AAG71509	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1190.	1510	96
477	gi2792016	Homo sapiens	olfactory receptor	1388	99
477	gi4092819	Homo sapiens	BC319430_5	1381	99
478	AAY73483	Homo sapiens	GEMY Human secreted protein clone yl18_1 protein sequence SEQ ID NO:188.	579	47
478	AAM92890	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 2239.	384	52
478	AAU83621	Homo sapiens	GETH Human PRO protein, Seq ID No 60.	333	28
479	AAM93439	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3078.	1182	94
479	gi15079907	Homo sapiens	Similar to secretory carrier membrane protein 4	1182	94
479	ABB06156	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:248.	1020	83
480	gi1497861	fowl adenovirus 8 [Fowl adenovirus 8]	fiber	81	24
480	gi6572647	fowl adenovirus 8	short fiber homolog [Fowl	81	24
480	gi3808227	Sphaeropsis sapinea RNA virus 2	coat protein	79	32
481	gi13517508	Homo sapiens	sodium bicarbonate cotransporter	5138	100
481	gi14582760	Homo sapiens	anion exchanger AE4	4979	97

173

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
481	gi7363254	Homo sapiens	sodium bicarbonate cotransporter 5	4973	97
482	AAM50714	Homo sapiens	MILL- Human TRP-like calcium channel-4 (TLCC-4).	2810	99
482	gi21435923	Homo sapiens	cation channel TRPV3	2810	99
482	gi20908451	Mus musculus	TRP ion channel TRPV3	2665	94
483	AAB86365	Homo sapiens	MEMO- Human ceramidase K3 protein.	1069	76
483	gi17529684	Mus musculus	cancer related gene-liver 1	1020	70
483	gi18028135	Drosophila melanogaster	brain washing	442	36
484	ABB89360	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1736.	251	78
484	gi1574439	Haemophilus influenzae Rd	leucine responsive regulatory protein (lrp)	73	38
484	gi12720483	Pasteurella multocida	Lrp	73	38
485	AAU99347	Homo sapiens	GETH Human PRO1113 (UNQ556) amino acid sequence SEQ ID NO:24.	2250	99
485	gi15987499	Mus musculus	tumor endothelial marker 5 precursor	1863	48
485	AAU74824	Homo sapiens	INCY- Human REPTR 7 protein.	1812	47
486	AAS12581_aa1	Homo sapiens	PEKE cDNA encoding novel human G protein-coupled receptor (GPCR).	1853	100
486	AAS07946_aa1	Homo sapiens	AREN- Human cDNA encoding G-protein coupled receptor, hRUP19.	1853	100
486	AAD27497_aa1	Homo sapiens	EURO- Human G-protein coupled receptor (GPCRx14) DNA.	1853	100
487	gi4959568	Homo sapiens	nuclear pore complex interacting protein NPIP	1087	67
487	ABB90262	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2638.	852	71
487	gi14603481	Homo sapiens	Similar to nuclear pore complex interacting protein	644	82
488	AAM25630	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1145.	554	90
488	AAG63804	Homo sapiens	NISC- Amino acid sequence of a human amino acid transporter.	551	98
488	gi9309293	Homo sapiens	asc-type amino acid transporter 1	551	98
489	AAM39751	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2896.	2304	99
489	AAM41538	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6469.	2294	99
489	AAM41537	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6468.	2294	99
490	AAE06056	Homo sapiens	HUMA- Human gene 16 encoded secreted protein HMIAP86, SEQ ID NO:118.	1006	75
490	AAU87079	Homo	HUMA- Human secreted protein sequence	1006	75

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	SEQ ID NO:118.		
490	AAY78511	Homo sapiens	AMYL- Human uncoupling protein 4 (UCP-4) amino acid sequence.	1006	75
491	AAG71803	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1484.	1616	100
491	ABB06625	Homo sapiens	CURA- G protein-coupled receptor GPCR13 protein SEQ ID NO:60.	1608	99
491	ABB06626	Homo sapiens	CURA- G protein-coupled receptor GPCR13b protein SEQ ID NO:62.	1605	99
492	gi10440458	Homo sapiens	FLJ00065 protein	992	100
492	gi15545993	Homo sapiens	Bcl-2 modifying factor	992	100
492	gi15545991	Mus musculus	Bcl-2 modifying factor	864	87
493	AAG67525	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	1841	99
493	ABB90207	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2583.	557	38
493	AAB69185	Homo sapiens	SREN- Human hSLR-iso protein SEQ ID NO:7.	557	38
494	ABB05727	Homo sapiens	GEHU- Human signal transduction protein clone tes3 5k22.	777	46
494	AAB12529	Homo sapiens	SLOK Human Ma5 protein SEQ ID NO:13.	777	46
494	gi6179740	Homo sapiens	paraneoplastic neuronal antigen MA3	777	46
495	gi17862902	Drosophila melanogaster	SD02518p	845	43
495	gi17861532	Drosophila melanogaster	GH11618p	833	42
495	gi530088	Glycine max	aminoalcoholphosphotransferase	398	28
496	gi9963853	Homo sapiens	HT018	1368	100
497	ABB90073	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2449.	1286	70
497	AAB12123	Homo sapiens	PROT- Hydrophobic domain protein from clone HP10608 isolated from Saos-2 cells.	1286	70
497	gi13241761	Homo sapiens	transmembrane protein induced by tumor necrosis factor alpha	1286	70
498	ABB85001	Homo sapiens	GETH Human PRO28631 protein sequence SEQ ID NO:370.	131	27
498	AAY86234	Homo sapiens	HUMA- Human secreted protein HNTNC20, SEQ ID NO:149.	123	38
498	AAB65258	Homo sapiens	GETH Human PRO1153 (UNQ583) protein sequence SEQ ID NO:351.	111	54
499	AAB93704	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13287.	3677	99
499	ABB07504	Homo sapiens	INCY- Human GTP-binding protein (GTPB) (ID: 4028409CD1).	2960	57
499	ABB07686	Homo sapiens	MERE Human GTPase-like protein, MFQ-111.	2456	56
500	gi21212948	Mus	peroxisomal protein (PeP)	462	53

175  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		musculus			
500	gi310897	Thermobifida fusca	beta-1,4-endoglucanase precursor	124	35
500	gi485747	Gallus gallus	protein-tyrosine phosphatase	115	32
501	AAB35156	Homo sapiens	SMIK Human nuclear receptor NOT1a splice variant related protein.	2750	88
501	AAU09156	Homo sapiens	SMIK Human NOT1 orphan nuclear receptor.	2750	88
501	AAR48631	Homo sapiens	MAGE/ Sequence of nuclear receptor of T-cells (NPT) steroidreceptor protein.	2750	88
502	AAU11383	Homo sapiens	SENO- Human T2R55 (hT2R55) polypeptide.	1632	98
502	gi20336515	Homo sapiens	candidate taste receptor T2RP24	1632	98
502	AAU11382	Homo sapiens	SENO- Human T2R54 (hT2R54) polypeptide.	894	57
503	AAB92909	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11539.	3006	98
503	gi17862912	Drosophila melanogaster	SD02996p	1037	31
503	ABB90736	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 204.	410	24
504	ABB05730	Homo sapiens	ZYMO Human zcytor17 protein sequence SEQ ID NO:2.	3070	99
504	gi20563277	Homo sapiens	gp130-like monocyte receptor	3070	99
504	ABB05741	Homo sapiens	ZYMO Human zcytor17 protein sequence SEQ ID NO:54.	3066	99
505	AAU80509	Homo sapiens	INCY- Human G-coupled receptor (GCREC) protein, Seq ID No 17.	1781	100
505	AAU11885	Homo sapiens	CURA- Human novel G protein-coupled receptor, GPCR1a.	1595	100
505	AAU11886	Homo sapiens	CURA- Human novel G protein-coupled receptor, GPCR1b.	1589	99
506	gi4102877	Mus musculus	Shc binding protein	2283	69
506	gi12017952	Homo sapiens	GE36	464	30
506	gi20906085	Methanosarcina mazei Goel	surface layer protein B	128	23
507	AAB11699	Homo sapiens	FUSO Human serine protease BSSP2 (hBSSP2), SEQ ID NO:10.	1404	100
507	gi12248917	Homo sapiens	spinesin	1404	100
507	AAE14342	Homo sapiens	INCY- Human protease PRTS-7 protein.	1236	99
508	gi18032273	Mus musculus	VPS10 domain receptor SorCS1c splice variant	5198	96
508	gi18032275	Homo sapiens	VPS10 domain receptor SorCS	5121	99
508	gi7715916	Mus musculus	SorCSb splice variant of the VPS10 domain receptor SorCS	4963	96

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
509	gi14278927	Mus musculus	gliacolin	1291	94
509	gi10566471	Mus musculus	Gliacolin	1291	94
509	gi3747097	Homo sapiens	Clq-related factor	976	70
510	gi12247892	Sterkiella histriomuscorum	SPEC3-like protein	90	31
510	AAA99908_aal	Homo sapiens	GETH cDNA encoding human protein PRO321.	71	30
510	ABB84833	Homo sapiens	GETH Human PRO321 protein sequence SEQ ID NO:34.	71	30
511	ABB90246	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2622.	648	100
511	AAB25755	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 33 SEQ ID NO:144.	648	100
511	AAB25754	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 33 SEQ ID NO:143.	301	100
512	gi13810306	Homo sapiens	transmembrane protein 7	1271	100
512	gi18250724	Mus musculus	transmembrane protein 7	639	64
512	gi15341942	Homo sapiens	28kD interferon responsive protein	428	38
513	AAG72504	Homo sapiens	YEDA Human OR-like polypeptide query sequence, SEQ ID NO: 2185.	1615	99
513	AAU24651	Homo sapiens	SENO- Human olfactory receptor AOLFR147.	1615	99
513	AAG71709	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1390.	1611	99
514	gi20381191	Homo sapiens	Similar to RIKEN cDNA 4932443L08 gene	2831	99
514	AAB83079	Homo sapiens	SMIK Human CASB6411 protein.	1806	100
514	AAB08764	Homo sapiens	INCY- A human leukocyte and blood related protein (LBAP).	1424	100
515	gi20072886	Homo sapiens	Similar to RIKEN cDNA 2610024A01 gene	1456	100
515	AAB74716	Homo sapiens	INCY- Human membrane associated protein MEMAP-22.	1094	99
515	ABB89524	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1900.	513	98
516	AAG66141	Homo sapiens	MILL- Human LGR6 polypeptide (clone Fbh150881).	3804	99
516	AAG66140	Homo sapiens	MILL- Human LGR6 polypeptide (clone fahr).	3804	99
516	gi10441732	Homo sapiens	leucine-rich repeat-containing G protein-coupled receptor 6	3782	100
517	AAB24465	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 29 SEQ ID NO:90.	447	98
518	AAM40227	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3372.	909	34
518	gi21321124	Rattus norvegicus	proton-associated sugar transporter A	898	34

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
518	gi4680229	Homo sapiens	DNb-5	537	29
519	ABB07253	Homo sapiens	LEXI- Human novel GPCR (NGPCR) protein.	3943	99
519	AAM69607	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29913.	1770	82
519	AAM57201	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29306.	1770	82
520	AAM43601	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 279.	1229	99
520	AAU18290	Homo sapiens	HUMA- Human endocrine polypeptide SEQ ID No 245.	1228	99
520	AAY27577	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 11.	598	100
521	AAB94304	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14767.	1523	100
521	AAD23974_aa1	Homo sapiens	INCY- Human neurotransmitter transporter, NTT-2 cDNA.	1350	92
521	AAE14404	Homo sapiens	INCY- Human neurotransmitter transporter, NTT-2.	1350	92
522	AAB74730	Homo sapiens	INCY- Human membrane associated protein MEMAP-36.	637	37
522	AAY94906	Homo sapiens	GEMY Human secreted protein clone rb649_3 protein sequence SEQ ID NO:18.	637	37
522	AAM40237	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3382.	523	37
523	AAB43665	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1110.	1254	100
523	AAY19759	Homo sapiens	HUMA- SEQ ID NO 477 from WO9922243.	966	100
523	gi21428606	Drosophila melanogaster	LD47425p	939	70
524	AAH42183_aa2	Homo sapiens	PHAA Nucleotide sequence of a G-protein coupled receptor.	1925	94
524	ABB06303	Homo sapiens	TAKE Human ZAQ protein sequence SEQ ID NO:1.	1925	94
524	AAB70143	Homo sapiens	TAKE Human G protein-coupled receptor protein.	1925	94
525	AAB93258	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12282.	930	53
525	AAY28810	Homo sapiens	GEMY nn296_2 secreted protein.	930	53
525	gi17944467	Drosophila melanogaster	RH03777p	749	48
526	AAM48989	Homo sapiens	TAKE Human testis originated G-protein coupled receptor TGR10.	1061	97
526	gi13876663	lumpy skin disease virus	G-protein-coupled chemokine receptor-like protein	191	25
526	gi7108517	Oryctolagus cuniculus	chemokine receptor	190	29
527	gi12214288	Homo sapiens	dJ402H5.2 (novel protein similar to worm and fly proteins)	2655	100
527	gi3880799	Caenorhabditis	Y39A1B.2	431	23

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		tis elegans			
527	gi15718594	Caenorhabditis elegans	C. elegans PTR-10 protein (corresponding sequence F55F8.1)	430	23
528	ABB89636	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2012.	817	100
528	gi21483396	Drosophila melanogaster	LD22376p	813	40
528	gi18480372	Mus musculus	olfactory receptor MOR145-3	82	25
529	AAM50125	Homo sapiens	MILL- Human acyltransferase 46743.	1874	100
529	AAB65222	Homo sapiens	GETH Human PRO1108 (UNQ551) protein sequence SEQ ID NO:248.	1583	69
529	AAM00959	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 435.	1583	69
530	ABB11531	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1901.	1290	99
530	AAM25596	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1111.	1289	99
530	ABB55767	Homo sapiens	FECH/ Human polypeptide SEQ ID NO 140.	1282	99
531	AAI66039_aal	Homo sapiens	KYOW Human G protein-coupled receptor encoding cDNA SEQ ID NO 2.	787	100
531	AAA64346_aal	Homo sapiens	MILL- DNA encoding a human G-protein coupled receptor designated 14273.	787	100
531	AAE04564	Homo sapiens	INCY- Human G-protein coupled receptor-20 (GCREC-20) protein.	787	100
532	AAU11888	Homo sapiens	CURA- Human novel G protein-coupled receptor, GPCR3a.	1747	99
532	AAU24662	Homo sapiens	SENO- Human olfactory receptor AOLFR160.	1747	99
532	AAU11889	Homo sapiens	CURA- Human novel G protein-coupled receptor, GPCR3b.	1632	98
533	gi557822	Saccharomyces cerevisiae	mal5, stal, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	314	25
533	gi1304387	Saccharomyces cerevisiae var. diastaticus	glucoamylase	314	25
533	gi915208	Sus scrofa	gastric mucin	307	25
534	AAU00437	Homo sapiens	COUN- Human dendritic cell membrane protein FIRE.	1997	88
534	AAV91625	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	1836	96
534	gi16930385	Mus musculus	seven-span membrane protein FIRE	1445	62
535	AAB61148	Homo sapiens	CURA- Human NOV17 protein.	2306	59
535	gi18676416	Homo sapiens	FLJ00080 protein	1900	57
535	AAB61147	Homo sapiens	CURA- Human NOV16 protein.	1378	53

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
536	AAB61148	Homo sapiens	CURA- Human NOV17 protein.	2306	59
536	gi18676416	Homo sapiens	FLJ00080 protein	1900	57
536	AAB61147	Homo sapiens	CURA- Human NOV16 protein.	1378	53
537	gi14325132	Thermoplasma volcanium	tricorn protease	75	29
537	gi21064441	Drosophila melanogaster	RE29777p	74	30
537	gi 13541726 ref NP_111414.1	Thermoplasma volcanium	Tricorn protease	75	29
538	AAG71899	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1580.	1603	100
538	AAU24548	Homo sapiens	SENO- Human olfactory receptor AOLFR35.	1603	100
538	AAE06770	Homo sapiens	INCY- Human G-protein coupled receptor-20 (GCREC-20) protein.	1598	100
539	AAG81420	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:358.	403	98
539	AAM93259	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2709.	327	38
539	gi16877659	Homo sapiens	Similar to RIKEN cDNA 1810054O13 gene	314	38
540	AAG89209	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 329.	460	97
540	gi1890812	Flexamia graminea	NADH dehydrogenase 1	71	24
540	gi 21295981 gb EAA08126.1	Anopheles gambiae str. PEST	agCP1281	73	28
541	ABB89210	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1586.	851	99
541	AA Y73442	Homo sapiens	GEMY Human secreted protein clone ya66 1 protein sequence SEQ ID NO:106.	596	95
541	AAB63255	Homo sapiens	LUDW- Human breast cancer associated antigen protein sequence SEQ ID NO:617.	88	40
542	gi9929918	Homo sapiens	intestinal mucin	4024	99
542	gi11990203	Homo sapiens	MUC3B mucin	3985	98
542	gi9929920	Homo sapiens	intestinal mucin	3908	96
543	gi17483744	Mus musculus	RING finger protein 33	1115	47
543	gi14043332	Homo sapiens	Similar to ring finger protein 23	913	40
543	gi10716078	Mus musculus	testis-abundant finger protein	907	40
544	AAG76127	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6891.	260	68
544	AAG03891	Homo	GEST Human secreted protein, SEQ ID NO:	260	68



180  
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens	7972.		
544	gi57131	Rattus norvegicus	ribosomal protein S26	260	68
545	AAU74820	Homo sapiens	INCY- Human REPTR 3 protein.	1737	42
545	gi6683905	Drosophila melanogaster	Dispatched	1073	31
545	AAU03497	Homo sapiens	UYZU- Human sterol sensing domain protein.	885	43
546	AAM78329	Homo sapiens	HYSE- Human protein SEQ ID NO 991.	933	70
546	ABL41227_aal	Homo sapiens	SWIT- Human G-protein coupled receptor encoding cDNA SEQ ID NO 8.	585	58
546	AAS16914_aal	Homo sapiens	PEKE Human G-protein coupled receptor (GPCR) cDNA.	585	58
547	gi20067221	Homo sapiens	Down syndrome cell adhesion molecule 2	11077	100
547	gi18033452	Homo sapiens	Down syndrome cell adhesion molecule DSCAML1	10745	99
547	AAM39040	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2185.	9116	100
548	gi12656633	Homo sapiens	transmembrane gamma-carboxyglutamic acid protein 3 TMG3	1192	100
548	AAM93243	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2675.	1186	99
548	gi20977032	Xenopus laevis	mitotic phosphoprotein 77	359	38
549	AAG89138	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 258.	709	74
549	AAE13062	Homo sapiens	AMGE- Human CD20/IgE-receptor like protein, agp-96614-al.	709	74
549	gi11559214	Homo sapiens	MS4A5	709	74
550	AAG72074	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1755.	1853	100
550	AAG71493	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1174.	1853	100
550	gi12054409	Homo sapiens	olfactory receptor	1853	100
551	AAB47932	Homo sapiens	SEIN/ Human Na <sup>+</sup> -driven Cl <sup>-</sup> /HCO <sub>3</sub> <sup>-</sup> -exchanger.	5677	99
551	gi11275360	Homo sapiens	NCBE	5677	99
551	gi11182364	Mus musculus	NCBE	5542	96
552	AAE04178	Homo sapiens	HUMA- Human gene 3 encoded secreted protein fragment, SEQ ID NO:169.	1111	98
552	AAE04127	Homo sapiens	HUMA- Human gene 3 encoded secreted protein HSDJL42, SEQ ID NO:114.	1078	98
552	AAE04102	Homo sapiens	HUMA- Human gene 3 encoded secreted protein HSDJL42, SEQ ID NO:88.	1068	98

181

Table 3

SEQ ID NO:	Database entry ID	Description	Results*
277	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 3.753e-10 235-250
278	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 3.753e-10 211-226
281	PD01572	PHOTOSYSTEM II REACTION CENTRE T PROTEIN PHOTOS.	PD01572 8.77 4.083e-09 1-30
282	BL00421	Transmembrane 4 family proteins.	BL00421E 20.97 4.000e-20 137-166 BL00421C 12.89 6.571e-12 77-88 BL00421A 11.79 1.563e-11 7-25
282	PR00259	TRANSMEMBRANE FOUR FAMILY SIGNATURE	PR00259D 13.50 8.200e-12 140-166 PR00259C 16.40 1.684e-09 13-41 PR00259A 9.27 4.405e-09 11-34
282	PR00218	PERIPHERIN (RDS)/ROM-1 FAMILY SIGNATURE	PR00218D 6.22 4.894e-09 76-104
286	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 5.355e-09 373-397
290	PR00970	ARGININE ADP-RIBOSYLTRANSFERASE SIGNATURE	PR00970A 17.73 6.906e-21 30-51 PR00970D 9.96 8.920e-20 133-149 PR00970F 12.30 9.250e-15 199-215 PR00970E 11.23 1.265e-14 178-193 PR00970G 9.97 3.700e-14 220-235 PR00970C 11.05 7.000e-14 90-104 PR00970B 16.37 7.387e-13 59-77
290	BL01291	NAD:arginine ADP-ribosyltransferases proteins.	BL01291F 23.30 5.974e-40 180-232 BL01291D 19.99 9.471e-31 115-148 BL01291A 22.07 4.892e-26 29-58 BL01291C 14.06 7.387e-17 87-102 BL01291G 15.18 4.176e-16 243-261 BL01291B 9.15 2.800e-11 69-82 BL01291E 7.03 1.000e-09 161-170
292	BL00983	Ly-6 / u-PAR domain proteins.	BL00983C 12.69 4.326e-10 92-107
292	BL00272	Snake toxins proteins.	BL00272C 8.27 9.372e-09 96-107
294	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290B 13.17 9.308e-15 168-185 BL00290A 20.89 1.450e-12 129-151
295	BL00571	Amidases proteins.	BL00571 25.69 4.188e-31 195-246
296	BL01271	Sodium:sulfate symporter family proteins.	BL01271D 25.26 1.000e-40 505-559 BL01271C 13.62 6.824e-21 432-453 BL01271B 12.02 9.206e-21 240-264 BL01271A 8.06 8.800e-20 131-150
298	PD00131	ATP-BINDING TRANSPORT TRANSMEMBR.	PD00131B 34.97 9.308e-32 480-533 PD00131C 19.59 1.000e-29 628-665
298	BL00211	ABC transporters family proteins.	BL00211B 13.37 7.750e-29 580-611 BL00211A 12.23 2.588e-10 474-485
298	PR00988	URIDINE KINASE SIGNATURE	PR00988A 6.39 6.838e-09 469-486
304	PD01572	PHOTOSYSTEM II REACTION CENTRE T PROTEIN PHOTOS.	PD01572 8.77 4.083e-09 1-30
308	BL00942	glpT family of transporters proteins.	BL00942B 20.36 1.750e-10 82-124 BL00942F 15.07 1.771e-10 339-356 BL00942C 14.04 6.610e-09 171-190
308	PD02963	COMPONENT PHOSPHOTRANSFERASE SYST.	PD02963B 5.41 6.776e-09 342-357
309	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 5.909e-21 59-80
309	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 9.743e-13 90-129
309	PR00237	RHODOPSIN-LIKE GPCR	PR00237B 13.50 9.280e-12 59-80

182  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
		SUPERFAMILY SIGNATURE	PR00237C 15.69 6.914e-10 104-126 PR00237A 11.48 4.774e-09 26-50
311	PR00254	NICOTINIC ACETYLCHOLINE RECEPTOR SIGNATURE	PR00254A 11.23 5.765e-14 64-80 PR00254D 15.50 2.023e-12 134-152 PR00254B 12.97 1.973e-11 98-112
311	BL00236	Neurotransmitter-gated ion-channels proteins.	BL00236A 21.96 5.050e-25 57-94 BL00236C 25.16 7.097e-25 139-177 BL00236D 25.66 8.105e-21 223-264 BL00236B 14.67 3.813e-11 111-120
311	PR00252	NEUROTRANSMITTER-GATED ION CHANNEL FAMILY SIGNATURE	PR00252A 14.28 5.696e-14 77-93 PR00252C 17.49 9.775e-12 154-168 PR00252B 15.17 2.406e-10 110-121
312	PD02327	GLYCOPROTEIN ANTIGEN PRECURSOR IMMUNOGLO.	PD02327B 19.84 2.091e-09 144-165
312	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 7.652e-09 291-300
313	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 8.043e-10 164-177 PR00019B 11.36 7.120e-09 136-149
313	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 7.319e-09 319-342
316	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 2.600e-10 45-84
316	PR00534	MELANOCORTIN RECEPTOR FAMILY SIGNATURE	PR00534A 11.49 9.446e-10 6-18
316	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 4.750e-18 193-208 PR00245A 18.03 4.808e-15 14-35 PR00245E 12.40 9.043e-11 246-260 PR00245B 10.38 2.102e-09 132-146
316	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 8.875e-09 59-81
320	PR00518	5-HYDROXYTRYPTAMINE 5A RECEPTOR SIGNATURE	PR00518D 8.59 9.471e-21 230-246 PR00518E 11.20 8.898e-12 246-255 PR00518C 5.94 1.000e-11 180-188
320	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 4.462e-19 118-140 PR00237G 19.63 7.261e-16 317-343 PR00237F 13.57 1.857e-15 280-304 PR00237E 13.03 4.600e-14 198-221 PR00237D 8.94 1.900e-11 154-175 PR00237B 13.50 7.517e-11 72-93
320	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 4.938e-27 104-143 BL00237C 13.19 2.500e-17 275-301 BL00237D 11.23 5.846e-11 327-343 BL00237B 5.28 6.727e-09 206-217
321	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 8.714e-12 17-41 PR00237G 19.63 4.600e-11 291-317 PR00237B 13.50 3.531e-10 50-71
326	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 6.657e-15 152-171 PR00007C 15.60 2.047e-14 200-221 PR00007A 19.33 8.412e-12 125-151
326	BL00415	Synapsins proteins.	BL00415N 4.29 7.307e-09 63-106
326	BL01113	C1q domain proteins.	BL01113B 18.26 3.647e-27 131-166 BL01113A 17.99 1.000e-13 68-94 BL01113C 13.18 2.532e-13 200-219 BL01113A 17.99 7.081e-13 59-85 BL01113A 17.99 8.297e-13 56-82 BL01113A 17.99 3.538e-12 65-91

Table 3

SEQ ID NO:	Database entry ID	Description	Results*
			BL01113A 17.99 5.385e-12 71-97 BL01113A 17.99 5.909e-11 74-100 BL01113A 17.99 8.773e-11 62-88 BL01113A 17.99 9.135e-09 53-79
326	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 4.808e-12 56-84 BL00420A 20.42 8.967e-10 53-81 BL00420A 20.42 7.231e-09 71-99 BL00420A 20.42 9.169e-09 77-105
330	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237E 13.03 6.400e-12 76-99 PR00237D 8.94 1.450e-11 26-47
330	BL00237	G-protein coupled receptors proteins.	BL00237C 13.19 7.000e-09 114-140 BL00237B 5.28 9.182e-09 84-95
333	BL00943	Cytochrome c oxidase assembly factor COX10/ctaB/cyoE signatur.	BL00943A 22.06 6.087e-17 117-155
334	PD00866	GLYCOPROTEIN PROTEIN SPIKE E2 PRECURSOR PEPLIMER.	PD00866L 3.73 6.902e-09 172-181
338	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 5.371e-10 103-125
338	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 2.473e-14 58-79 PR00245B 10.38 5.500e-13 176-190 PR00245E 12.40 2.149e-11 290-304 PR00245D 10.47 5.814e-10 273-284
338	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 4.818e-14 89-128 BL00237D 11.23 5.364e-09 281-297
339	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 5.371e-10 103-125
339	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 2.473e-14 58-79 PR00245B 10.38 5.500e-13 176-190 PR00245D 10.47 5.814e-10 273-284
339	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 4.818e-14 89-128 BL00237D 11.23 5.364e-09 281-297
340	PR00878	CHOLINESTERASE SIGNATURE	PR00878F 5.37 4.780e-13 523-535
340	BL00122	Carboxylesterases type-B serine proteins.	BL00122E 22.02 1.563e-25 254-294 BL00122A 12.04 5.929e-16 69-89 BL00122D 12.53 4.484e-14 230-245 BL00122B 16.84 5.800e-14 139-149 BL00122G 11.67 8.615e-13 561-571 BL00122C 7.91 3.118e-11 201-211 BL00122F 11.10 3.000e-10 306-315
340	BL01173	Lipolytic enzymes G-D-X-G family, histidine.	BL01173A 9.41 5.245e-10 203-215
341	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.564e-13 711-736
341	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 4.323e-10 713-736
341	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 9.775e-09 122-137
342	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 5.629e-13 90-129
342	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 2.565e-17 59-80 PR00245E 12.40 9.735e-13 226-240 PR00245C 7.84 3.591e-09 174-189
343	PF00954	S-locus glycoprotein family.	PF00954E 23.75 6.798e-09 152-202
343	BL00246	Wnt-1 family proteins.	BL00246E 20.32 8.306e-09 141-186
344	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 9.455e-14 93-132
344	PR00245	OLFACTORY RECEPTOR	PR00245A 18.03 1.000e-18 62-83

184  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
		SIGNATURE	PR00245B 10.38 9.143e-16 180-194 PR00245C 7.84 1.360e-13 241-256 PR00245E 12.40 7.882e-13 294-308 PR00245D 10.47 1.000e-10 277-288
344	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 4.600e-10 107-129 PR00237G 19.63 1.209e-09 275-301
345	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 9.129e-11 464-487 PR00249E 14.90 4.493e-10 549-574
345	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.073e-13 462-487 BL00649E 15.34 2.857e-12 549-578 BL00649G 13.52 8.826e-11 722-747 BL00649B 20.68 8.548e-09 406-451
345	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 7.600e-11 87-102 BL01187A 9.98 1.000e-08 68-79
346	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 9.129e-11 368-391 PR00249E 14.90 4.493e-10 453-478
346	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.073e-13 366-391 BL00649E 15.34 2.857e-12 453-482 BL00649G 13.52 8.826e-11 626-651 BL00649B 20.68 8.548e-09 310-355
355	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 9.500e-11 144-157 PR00019A 11.19 5.696e-10 147-160 PR00019B 11.36 6.400e-10 95-108 PR00019B 11.36 5.320e-09 119-132
355	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014C 15.44 8.043e-09 435-453
357	BL00427	Disintegrins proteins.	BL00427 13.93 9.384e-24 443-497
357	PR00289	DISINTEGRIN SIGNATURE	PR00289A 13.62 4.000e-14 457-476 PR00289B 11.79 6.745e-11 486-498
357	BL00142	Neutral zinc metalloproteases, zinc-binding region proteins.	BL00142 8.38 2.125e-10 343-353
358	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270C 19.54 4.919e-14 116-144 PD01270B 22.18 4.462e-10 73-109
359	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270C 19.54 4.919e-14 110-138 PD01270B 22.18 4.462e-10 67-103
368	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463E 17.37 4.667e-12 344-370
368	PR00385	P450 SUPERFAMILY SIGNATURE	PR00385A 14.97 1.783e-13 335-352 PR00385B 10.22 5.950e-12 353-366
368	PR00464	E-CLASS P450 GROUP II SIGNATURE	PR00464C 18.84 7.750e-22 324-352 PR00464A 20.47 7.300e-17 149-169 PR00464D 17.40 6.538e-14 353-370 PR00464B 20.41 1.000e-11 205-223
368	PR00408	MITOCHONDRIAL P450 SIGNATURE	PR00408D 15.44 8.099e-09 335-352
370	PR00001	COAGULATION FACTOR GLA DOMAIN SIGNATURE	PR00001B 10.75 9.000e-15 70-83 PR00001A 12.78 5.800e-10 56-69
371	BL00406	Actins proteins.	BL00406D 12.58 3.143e-19 257-311 BL00406A 9.95 5.729e-13 15-49 BL00406B 5.47 7.429e-12 51-105 BL00406C 6.75 9.682e-12 110-164
371	PR00735	GLYCOSYL HYDROLASE FAMILY 8 SIGNATURE	PR00735D 12.75 1.000e-08 363-374
377	BL00120	Lipases, serine proteins.	BL00120B 11.37 1.383e-10 124-138
377	PR00793	PROLYL AMINOPEPTIDASE (S33)	PR00793C 12.24 9.500e-09 128-142

185  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
		FAMILY SIGNATURE	
378	BL00120	Lipases, serine proteins.	BL00120B 11.37 1.383e-10 124-138
378	PR00793	PROLYL AMINOPEPTIDASE (S33) FAMILY SIGNATURE	PR00793C 12.24 9.500e-09 128-142
382	PR00761	BINDIN PRECURSOR SIGNATURE	PR00761E 14.32 1.663e-09 188-206
388	PR00420	AROMATIC-RING HYDROXYLASE (FLAVOPROTEIN MONOOXYGENASE) SIGNATURE	PR00420A 14.78 4.638e-13 15-37
388	PR00757	FLAVIN-CONTAINING AMINE OXIDASE SIGNATURE	PR00757A 6.64 1.414e-10 15-34
388	PR00419	ADRENODOXIN REDUCTASE FAMILY SIGNATURE	PR00419A 14.89 4.094e-10 15-37
388	PR00072	MALIC ENZYME SIGNATURE	PR00072F 8.87 5.922e-09 16-32
388	BL00623	GMC oxidoreductases proteins.	BL00623A 12.60 8.200e-09 15-33
388	PR00368	FAD-DEPENDENT PYRIDINE NUCLEOTIDE REDUCTASE SIGNATURE	PR00368A 17.76 9.839e-09 15-37
396	BL00031	Nuclear hormones receptors DNA- binding region proteins.	BL00031A 19.55 9.471e-34 102-134 BL00031B 22.25 2.216e-22 135-166
396	PR00398	STEROID HORMONE RECEPTOR SIGNATURE	PR00398A 14.44 3.328e-16 102-119 PR00398C 13.47 1.450e-10 143-161
396	PR00350	VITAMIN D RECEPTOR SIGNATURE	PR00350B 9.35 2.125e-12 119-138 PR00350F 8.61 4.385e-10 399-422 PR00350A 10.48 7.871e-09 102-118
396	PR00047	C4-TYPE STEROID RECEPTOR ZINC FINGER SIGNATURE	PR00047A 15.70 5.500e-19 102-118 PR00047B 7.63 4.522e-17 118-133 PR00047D 13.53 9.550e-10 158-166 PR00047C 5.40 8.788e-09 150-158
398	PD01672	+ TRANSPORT EXCHANGER NA H TRANS.	PD01672B 15.16 1.115e-24 125-173 PD01672D 10.50 5.275e-18 207-243 PD01672I 17.98 5.939e-16 402-448 PD01672G 15.27 1.600e-12 318-351 PD01672C 16.18 3.933e-12 172-206 PD01672H 22.99 4.949e-10 355-401
403	PD02797	HYDROLASE CELL WALL N- ACETYLMURAMOYL-L-AL.	PD02797D 19.90 9.032e-09 120-159
405	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 8.861e-09 77-91
411	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 2.575e-09 104-126
411	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 9.419e-15 90-129 BL00237D 11.23 5.636e-09 282-298
411	PR00896	VASOPRESSIN RECEPTOR SIGNATURE	PR00896B 9.01 7.577e-09 55-66
411	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 9.053e-19 238-253 PR00245A 18.03 7.907e-18 59-80 PR00245E 12.40 2.731e-14 291-305 PR00245D 10.47 8.531e-09 274-285
412	PR00646	RDC1 ORPHAN RECEPTOR SIGNATURE	PR00646I 10.54 1.110e-26 301-320 PR00646D 15.99 1.540e-26 85-103 PR00646G 14.95 1.281e-25 173-190 PR00646B 6.02 1.978e-25 21-40 PR00646A 16.77 9.438e-24 4-21 PR00646F 10.13 1.150e-23 156-173 PR00646C 18.45 1.170e-23 49-64

186

Table 3

SEQ ID NO:	Database entry ID	Description	Results*
			PR00646E 9.52 5.500e-23 127-144 PR00646H 6.32 1.101e-20 219-234
412	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 4.789e-24 92-131 BL00237C 13.19 9.280e-14 227-253 BL00237D 11.23 7.857e-13 289-305
412	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 8.800e-18 106-128 PR00237B 13.50 2.000e-15 61-82 PR00237G 19.63 2.800e-15 279-305 PR00237F 13.57 1.000e-14 232-256 PR00237E 13.03 4.333e-11 195-218 PR00237D 8.94 4.375e-10 142-163
412	PR00425	BRADYKININ RECEPTOR SIGNATURE	PR00425C 13.23 8.286e-10 92-111
412	PR00526	FORMYL-METHIONYL PEPTIDE RECEPTOR SIGNATURE	PR00526C 13.54 9.550e-10 100-117
412	PR00241	ANGIOTENSIN II RECEPTOR SIGNATURE	PR00241C 8.90 4.536e-09 115-122
413	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.438e-12 117-131
415	PR00120	H <sup>+</sup> -TRANSPORTING ATPASE (PROTON PUMP) SIGNATURE	PR00120C 9.90 5.800e-19 802-818
415	PR00121	SODIUM/POTASSIUM-TRANSPORTING ATPASE SIGNATURE	PR00121D 16.72 1.209e-28 455-476 PR00121I 15.47 2.500e-26 1037-1061 PR00121B 7.83 6.786e-26 218-238 PR00121G 6.89 8.875e-26 941-961 PR00121H 12.14 9.100e-26 1003-1023 PR00121F 6.70 4.214e-25 874-895 PR00121C 9.40 7.652e-23 382-404 PR00121E 13.97 1.563e-22 592-610 PR00121A 6.71 7.429e-19 191-205
415	BL00154	E1-E2 ATPases phosphorylation site proteins.	BL00154E 20.37 8.615e-38 680-720 BL00154B 15.44 2.800e-31 420-456 BL00154G 21.18 9.526e-30 825-858 BL00154F 8.23 6.400e-28 799-822 BL00154C 12.38 6.000e-23 458-476 BL00154A 11.86 9.500e-16 276-293 BL00154D 12.57 3.769e-13 595-605
415	PR00119	P-TYPE CATION-TRANSPORTING ATPASE SUPERFAMILY SIGNATURE	PR00119E 8.48 6.250e-25 802-821 PR00119B 13.94 2.800e-20 462-476 PR00119A 17.34 3.000e-15 302-316 PR00119D 9.56 3.571e-13 696-706 PR00119C 11.01 6.143e-13 674-685 PR00119F 11.81 7.750e-13 826-838
415	BL01228	Hypothetical cof family proteins.	BL01228D 17.44 6.250e-11 800-824
415	BL01047	Heavy-metal-associated domain proteins.	BL01047B 19.73 6.063e-10 808-828
418	BL00219	Anion exchangers family proteins.	BL00219K 12.73 9.883e-24 677-718 BL00219M 9.98 5.208e-23 762-807 BL00219H 10.06 5.034e-22 474-521 BL00219N 10.66 7.545e-22 808-851 BL00219B 14.47 6.104e-20 194-237 BL00219I 6.16 9.818e-17 587-640 BL00219G 12.86 9.697e-16 434-472 BL00219A 17.13 1.000e-15 65-96

187

Table 3

SEQ ID NO:	Database entry ID	Description	Results*
			BL00219F 10.52 8.024e-15 381-404 BL00219C 17.29 4.470e-14 239-277 BL00219O 14.02 1.000e-13 853-892 BL00219E 11.63 2.019e-10 341-380 BL00219L 18.71 3.560e-10 719-757
418	PR00165	ANION EXCHANGER SIGNATURE	PR00165B 15.26 1.549e-13 376-396 PR00165I 10.02 2.521e-13 675-694 PR00165E 8.63 8.859e-11 463-482 PR00165F 10.39 7.674e-10 495-513 PR00165G 11.41 8.180e-09 588-607
421	DM00099	4 kw A55R REDUCTASE TERMINAL DIHYDROPTERIDINE.	DM00099B 14.73 2.125e-09 455-464
421	PR00501	KELCH REPEAT SIGNATURE	PR00501B 18.88 8.342e-09 453-467
421	BL00292	Cyclins proteins.	BL00292B 20.31 1.000e-08 432-462
422	BL00599	Aminotransferases class-II pyridoxal-phosphate attachment sit.	BL00599B 18.93 7.894e-12 394-422
422	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320B 12.19 5.500e-09 85-99 PR00320C 13.01 6.400e-09 186-200 PR00320A 16.74 6.927e-09 85-99 PR00320A 16.74 8.024e-09 186-200
423	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 8.780e-09 862-894
423	PF00761	Polyomavirus coat protein.	PF00761A 12.61 8.925e-09 461-485
427	PR00902	VP6 BLUE-TONGUE VIRUS INNER CAPSID PROTEIN SIGNATURE	PR00902J 18.54 6.400e-09 271-292
428	PR00902	VP6 BLUE-TONGUE VIRUS INNER CAPSID PROTEIN SIGNATURE	PR00902J 18.54 6.400e-09 271-292
430	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 4.273e-15 118-148
430	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 9.426e-13 118-136
430	BL00240	Receptor tyrosine kinase class III proteins.	BL00240E 11.56 6.743e-09 104-141
432	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 6.333e-09 32-40
435	PR00625	DNAJ PROTEIN FAMILY SIGNATURE	PR00625D 11.93 9.077e-09 59-69
438	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 6.186e-09 460-492
448	BL00031	Nuclear hormones receptors DNA- binding region proteins.	BL00031A 19.55 5.320e-30 11-43 BL00031B 22.25 6.604e-16 27-58
448	PR00350	VITAMIN D RECEPTOR SIGNATURE	PR00350A 10.48 1.692e-16 11-27 PR00350F 8.61 6.400e-11 290-313 PR00350B 9.35 7.581e-11 28-47 PR00350E 11.55 9.693e-11 242-261
448	PR00047	C4-TYPE STEROID RECEPTOR ZINC FINGER SIGNATURE	PR00047A 15.70 2.200e-16 11-27 PR00047B 7.63 3.813e-16 27-42 PR00047C 5.40 5.000e-10 42-50 PR00047D 13.53 6.850e-10 50-58
448	PR00546	THYROID HORMONE RECEPTOR SIGNATURE	PR00546H 16.85 6.523e-09 169-188
448	PR00398	STEROID HORMONE RECEPTOR SIGNATURE	PR00398A 14.44 7.750e-14 11-28 PR00398C 13.47 4.857e-09 35-53 PR00398F 13.87 7.943e-09 150-169
449	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 2.473e-10 217-234 PR00205B 11.39 8.691e-10 321-338
449	BL00232	Cadherins extracellular repeat proteins	BL00232B 32.79 5.279e-20 219-266



Table 3

SEQ ID NO:	Database entry ID	Description	Results*
		domain proteins.	BL00232C 10.65 6.268e-12 217-234 BL00232C 10.65 9.308e-10 321-338
449	PR00291	SOYBEAN TRYPSIN INHIBITOR (KUNITZ-TYPE) SIGNATURE	PR00291A 19.85 9.366e-09 225-254
449	PR00649	GPR6 ORPHAN RECEPTOR SIGNATURE	PR00649B 8.21 1.000e-08 252-269
452	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306B 5.57 9.000e-09 52-62
457	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290B 13.17 7.750e-19 52-69
458	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 4.966e-13 59-80 PR00245B 10.38 8.875e-13 177-191
458	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 5.500e-12 90-129
458	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 2.688e-10 59-80 PR00237C 15.69 7.171e-10 104-126 PR00237A 11.48 2.161e-09 26-50
464	BL00427	Disintegrins proteins.	BL00427 13.93 7.592e-26 379-433
464	PR00138	MATRIXIN SIGNATURE	PR00138D 16.56 5.101e-11 278-303
464	BL00142	Neutral zinc metalloproteinases, zinc-binding region proteins.	BL00142 8.38 7.545e-11 278-288
464	PR00289	DISINTEGRIN SIGNATURE	PR00289A 13.62 2.500e-14 393-412 PR00289B 11.79 4.226e-10 422-434
464	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 8.909e-10 273-291
464	PR00907	THROMBOMODULIN SIGNATURE	PR00907E 11.70 3.647e-09 591-613
464	BL00546	Matrixins cysteine switch.	BL00546C 16.41 4.255e-09 272-303
464	BL00024	Hemopexin domain proteins.	BL00024D 17.28 5.596e-09 272-303
466	DM01206	CORONAVIRUS NUCLEOCAPSID PROTEIN.	DM01206B 10.69 1.000e-08 9-28
470	PR00211	GLUTELIN SIGNATURE	PR00211B 0.86 5.673e-10 522-542
470	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 8.607e-09 591-603
470	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 4.051e-09 522-554 DM00215 19.43 6.644e-09 512-544 DM00215 19.43 9.085e-09 531-563
474	PR00220	SYNAPTOPHYSIN/SYNAPTOPORIN FAMILY SIGNATURE	PR00220D 8.32 7.585e-26 131-154 PR00220C 11.05 4.477e-25 99-123 PR00220A 10.93 8.244e-24 36-58 PR00220E 3.46 6.932e-23 197-215
474	BL00604	Synaptophysin / synaptoporin proteins.	BL00604E 8.32 1.444e-23 182-223 BL00604B 9.95 1.329e-19 86-115 BL00604C 14.66 5.639e-12 116-147 BL00604D 12.28 5.410e-11 148-182
476	PR00785	NUCLEAR TRANSLOCATOR SIGNATURE	PR00785H 15.80 7.692e-09 151-167
477	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 7.300e-19 62-83 PR00245C 7.84 8.579e-19 241-256 PR00245D 10.47 4.000e-15 277-288 PR00245B 10.38 4.405e-12 180-194 PR00245E 12.40 1.509e-10 294-308
477	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 6.143e-13 93-132 BL00237D 11.23 5.091e-09 285-301
478	BL00297	Heat shock hsp70 proteins family proteins.	BL00297D 11.95 8.835e-09 86-125
481	BL00219	Anion exchangers family proteins.	BL00219E 11.63 4.838e-24 376-415 BL00219K 12.73 9.883e-24 715-756

189  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
			BL00219M 9.98 5.208e-23 800-845 BL00219H 10.06 5.034e-22 509-556 BL00219N 10.66 7.545e-22 846-889 BL00219B 14.47 6.104e-20 218-261 BL00219I 6.16 9.818e-17 625-678 BL00219G 12.86 9.697e-16 469-507 BL00219F 10.52 8.024e-15 416-439 BL00219C 17.29 4.470e-14 263-301 BL00219O 14.02 1.000e-13 891-930 BL00219L 18.71 9.422e-10 757-795
481	PR00165	ANION EXCHANGER SIGNATURE	PR00165A 9.84 8.000e-18 386-408 PR00165B 15.26 1.549e-13 411-431 PR00165I 10.02 2.521e-13 713-732 PR00165E 8.63 8.859e-11 498-517 PR00165F 10.39 7.674e-10 530-548 PR00165G 11.41 8.180e-09 626-645
486	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 2.552e-13 260-286 PR00237B 13.50 3.045e-13 50-71 PR00237F 13.57 1.000e-10 218-242 PR00237A 11.48 9.333e-10 17-41 PR00237C 15.69 2.800e-09 95-117
486	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 3.032e-15 81-120 BL00237C 13.19 2.324e-10 213-239 BL00237D 11.23 2.607e-10 270-286 BL00237B 5.28 7.136e-09 185-196
490	BL00215	Mitochondrial energy transfer proteins.	BL00215A 15.82 7.618e-14 67-91
491	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 8.364e-14 59-80 PR00245C 7.84 5.500e-12 237-252 PR00245B 10.38 4.600e-11 177-191 PR00245E 12.40 9.830e-10 290-304
491	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 3.605e-10 271-297 PR00237C 15.69 6.175e-09 104-126
491	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 5.371e-13 90-129 BL00237D 11.23 9.455e-09 281-297
493	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 4.150e-10 117-130 PR00019B 11.36 9.100e-10 141-154 PR00019A 11.19 8.000e-09 120-133
493	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR00500B 7.74 9.337e-09 225-245
495	BL00379	CDP-alcohol phosphatidyltransferases proteins.	BL00379 24.64 8.855e-16 104-140
500	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 9.550e-10 107-137
501	BL00031	Nuclear hormones receptors DNA-binding region proteins.	BL00031B 22.25 6.538e-34 277-308
501	PR00047	C4-TYPE STEROID RECEPTOR ZINC FINGER SIGNATURE	PR00047C 5.40 3.250e-14 292-300 PR00047D 13.53 3.250e-12 300-308
501	PR00398	STEROID HORMONE RECEPTOR SIGNATURE	PR00398C 13.47 5.299e-14 285-303 PR00398G 15.17 7.081e-09 388-408
504	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR00500A 5.70 8.768e-10 55-73
504	PD02382	RECEPTOR CHAIN PRECURSOR TRANSME.	PD02382B 4.60 3.100e-09 263-269
504	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 7.643e-09 535-565

190

Table 3

SEQ ID NO:	Database entry ID	Description	Results*
505	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 6.870e-24 101-122 PR00245C 7.84 2.421e-19 280-295 PR00245E 12.40 8.714e-16 333-347 PR00245D 10.47 6.786e-13 316-327 PR00245B 10.38 6.906e-13 219-233
505	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 8.839e-15 132-171 BL00237D 11.23 2.364e-09 324-340
505	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 1.750e-09 101-122 PR00237C 15.69 4.600e-09 146-168 PR00237A 11.48 5.065e-09 68-92 PR00237G 19.63 5.605e-09 314-340
505	PR00023	ZONA PELLUCIDA SPERM-BINDING PROTEIN SIGNATURE	PR00023E 22.27 9.813e-09 170-187
507	PR00722	CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1) SIGNATURE	PR00722A 12.27 4.960e-15 244-259 PR00722C 10.87 2.929e-14 509-521
507	BL00134	Serine proteases, trypsin family, histidine proteins.	BL00134B 15.99 3.571e-19 510-533 BL00134A 11.96 3.160e-17 243-259 BL00134C 13.45 3.250e-13 546-559
507	BL00495	Apple domain proteins.	BL00495N 11.04 4.729e-24 502-536 BL00495O 13.75 6.127e-15 537-565 BL00495M 8.50 6.400e-12 429-463
507	BL01253	Type I fibronectin domain proteins.	BL01253H 13.15 8.364e-19 528-562 BL01253G 11.34 1.574e-17 509-522 BL01253F 14.35 6.850e-14 465-503 BL01253E 16.01 8.861e-14 427-463 BL01253D 4.84 6.400e-10 243-256
507	BL00021	Kringle domain proteins.	BL00021D 24.56 8.500e-28 518-559 BL00021B 13.33 5.154e-15 243-260 BL00021C 22.21 6.943e-09 438-459
509	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 6.657e-15 246-265 PR00007C 15.60 2.047e-14 294-315 PR00007A 19.33 8.412e-12 219-245
509	BL00415	Synapsins proteins.	BL00415N 4.29 7.307e-09 157-200
509	BL01113	C1q domain proteins.	BL01113B 18.26 3.647e-27 225-260 BL01113A 17.99 1.000e-13 162-188 BL01113C 13.18 2.532e-13 294-313 BL01113A 17.99 7.081e-13 153-179 BL01113A 17.99 8.297e-13 150-176 BL01113A 17.99 3.538e-12 159-185 BL01113A 17.99 5.385e-12 165-191 BL01113A 17.99 5.909e-11 168-194 BL01113A 17.99 8.773e-11 156-182 BL01113A 17.99 9.135e-09 147-173
509	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 4.808e-12 150-178 BL00420A 20.42 8.967e-10 147-175 BL00420A 20.42 7.231e-09 165-193 BL00420A 20.42 9.169e-09 171-199
513	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 9.486e-13 92-131
513	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 6.714e-12 61-82 PR00245C 7.84 8.000e-10 240-255
513	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 5.355e-09 28-52 PR00237C 15.69 9.550e-09 106-128
516	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 2.543e-11 665-691 PR00237A 11.48 3.000e-10 419-443

191  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
516	PR00373	GLYCOPROTEIN HORMONE RECEPTOR SIGNATURE	PR00373D 11.16 2.403e-09 498-512
516	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 6.600e-10 491-530 BL00237D 11.23 4.545e-09 675-691
516	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 7.300e-11 210-223 PR00019A 11.19 8.043e-10 280-293 PR00019B 11.36 5.320e-09 207-220
516	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 7.429e-09 395-407
519	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.564e-13 578-603
519	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 4.323e-10 580-603
521	PR00176	SODIUM/NEUROTRANSMITTER SYMPORTER SIGNATURE	PR00176C 10.84 2.667e-24 142-168 PR00176A 16.82 5.500e-23 69-90 PR00176B 7.31 9.308e-17 98-117
521	BL00610	Sodium:neurotransmitter symporter family proteins.	BL00610A 17.73 1.000e-40 69-118 BL00610B 23.65 1.000e-40 133-182 BL00610C 12.94 6.157e-14 226-277
524	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 7.750e-14 93-114 PR00237C 15.69 1.667e-12 140-162 PR00237F 13.57 8.333e-12 278-302 PR00237E 13.03 6.667e-11 229-252 PR00237D 8.94 7.750e-10 174-195
524	BL00419	Photosystem I psaA and psaB proteins.	BL00419L 20.03 7.850e-09 11-59
524	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 3.739e-20 126-165 BL00237C 13.19 4.808e-13 273-299 BL00237B 5.28 8.773e-09 237-248
526	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237D 8.94 2.000e-09 171-192
526	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 3.020e-09 121-160
526	PR00641	EBI1 ORPHAN RECEPTOR SIGNATURE	PR00641E 10.22 8.975e-09 119-136
527	BL00519	Bacterial regulatory proteins, asnC family proteins.	BL00519C 29.50 6.595e-09 110-154
531	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 8.258e-15 143-182
531	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 7.375e-11 81-105 PR00237B 13.50 4.094e-10 113-134 PR00237C 15.69 2.575e-09 157-179
532	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 2.029e-13 111-150
532	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 9.000e-23 80-101 PR00245C 7.84 3.543e-14 259-274 PR00245B 10.38 9.357e-14 198-212 PR00245E 12.40 8.286e-12 312-326
532	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 2.161e-09 47-71 PR00237C 15.69 4.150e-09 125-147
533	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 1.000e-17 603-624
534	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 9.129e-11 247-270 PR00249E 14.90 4.493e-10 332-357
534	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.073e-13 245-270 BL00649E 15.34 2.857e-12 332-361 BL00649G 13.52 8.826e-11 505-530 BL00649B 20.68 8.548e-09 189-234
538	PR00245	OLFACTORY RECEPTOR	PR00245C 7.84 6.049e-15 238-253

192  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
		SIGNATURE	PR00245A 18.03 6.192e-15 59-80 PR00245E 12.40 4.643e-12 291-305 PR00245B 10.38 4.886e-10 177-191
538	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 5.500e-12 90-129 BL00237D 11.23 7.545e-09 282-298
538	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 2.674e-09 272-298 PR00237E 13.03 7.088e-09 199-222 PR00237C 15.69 8.875e-09 104-126
542	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 4.375e-10 411-436
542	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 3.508e-11 416-434 PR00011B 13.08 4.522e-10 416-434 PR00011A 14.06 2.479e-09 416-434
542	PR00962	LETHAL(2) GIANT LARVAE PROTEIN SIGNATURE	PR00962F 12.39 6.855e-09 517-536
543	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 4.857e-10 31-39
544	BL00733	Ribosomal protein S26e proteins.	BL00733A 11.62 8.784e-25 1-43 BL00733B 12.04 6.870e-20 44-76
544	BL00127	Pancreatic ribonuclease family proteins.	BL00127B 26.57 3.455e-09 134-178
546	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 8.313e-10 64-85 PR00237D 8.94 7.000e-09 145-166
547	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 7.480e-11 1216-1246 BL00790I 20.01 6.963e-10 1115-1145 BL00790I 20.01 8.988e-10 1314-1344 BL00790H 13.42 9.514e-10 1266-1291
547	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 1.305e-09 2034-2066
547	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 8.024e-12 1408-1440 PD02870D 15.74 9.900e-10 1408-1442 PD02870B 18.83 7.415e-09 339-371
547	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014A 8.22 3.864e-09 1265-1274 PR00014D 12.04 7.750e-09 1122-1136
547	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 8.043e-09 347-356
547	PD02327	GLYCOPROTEIN ANTIGEN PRECURSOR IMMUNOGLO.	PD02327B 19.84 9.591e-09 305-326 PD02327B 19.84 9.591e-09 676-697
547	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 7.907e-10 487-510 BL00240B 24.70 1.000e-08 305-328
548	PR00001	COAGULATION FACTOR GLA DOMAIN SIGNATURE	PR00001A 12.78 2.174e-13 23-36 PR00001B 10.75 8.364e-13 37-50 PR00001C 16.60 6.327e-09 51-65
550	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 2.500e-22 59-80 PR00245C 7.84 7.000e-18 238-253 PR00245B 10.38 7.480e-15 177-191 PR00245E 12.40 6.029e-13 291-305
550	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 6.182e-14 90-129 BL00237D 11.23 7.750e-10 282-298
550	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 5.219e-12 272-298 PR00237E 13.03 1.000e-10 199-222 PR00237C 15.69 3.925e-09 104-126
551	PR00165	ANION EXCHANGER SIGNATURE	PR00165A 9.84 1.652e-16 453-475

193  
Table 3

SEQ ID NO:	Database entry ID	Description	Results*
			PR00165B 15.26 7.835e-14 478-498 PR00165I 10.02 5.378e-12 781-800 PR00165D 7.84 8.159e-11 534-553 PR00165F 10.39 8.729e-11 597-615 PR00165H 8.01 1.321e-10 729-749
551	BL00219	Anion exchangers family proteins.	BL00219C 17.29 7.474e-25 338-376 BL00219N 10.66 4.575e-24 914-957 BL00219E 11.63 9.471e-24 443-482 BL00219K 12.73 2.098e-22 783-824 BL00219B 14.47 8.571e-22 293-336 BL00219M 9.98 7.222e-21 868-913 BL00219H 10.06 9.693e-21 576-623 BL00219A 17.13 4.176e-20 127-158 BL00219I 6.16 3.106e-19 693-746 BL00219L 18.71 3.889e-19 825-863 BL00219G 12.86 3.198e-17 536-574 BL00219F 10.52 7.152e-16 483-506 BL00219O 14.02 1.835e-11 959-998 BL00219D 15.15 3.148e-10 377-412

\*Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence.

Table 4A

SEQ ID NO:	Pfam Model	Description	E-value	Score
277	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	5.2e-10	36.7
278	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	5.2e-10	36.7
279	PA	PA domain	1.3e-18	75.3
282	transmembrane4	Tetraspanin family	1.7e-48	161.4
287	sushi	Sushi domain (SCR repeat)	1.8e-56	201.1
290	ART	NAD:arginine ADP-ribosyltransferase	6.5e-207	700.8
292	UPAR_LY6	u-PAR/Ly-6 domain	0.01	14.2
293	PMP22 Claudin	PMP-22/EMP/MP20/Claudin family	9.4e-06	32.5
294	MHC_II_alpha	Class II histocompatibility antigen, alpha domain	4.1e-44	160.0
295	Amidase	Amidase	4.6e-71	249.5
296	Na sulph symp	Sodium:sulfate symporter transmembrane region	1.3e-73	258.0
298	ABC membrane	ABC transporter transmembrane region.	1.6e-56	201.2
299	PMP22 Claudin	PMP-22/EMP/MP20/Claudin family	0.048	-29.1
306	Acyltransferase	Acyltransferase	9.6e-06	30.8
309	7tm_1	7 transmembrane receptor (rhodopsin family)	4.1e-30	97.8
311	Neur_chan_LBD	Neurotransmitter-gated ion-channel ligand binding domain	2.2e-83	290.4
312	ig	Immunoglobulin domain	4.7e-20	69.7
313	LRR	Leucine Rich Repeat	1.9e-23	91.3
314	Plexin repeat	Plexin repeat	0.02	20.2
315	Plexin repeat	Plexin repeat	0.02	20.2
316	7tm_1	7 transmembrane receptor (rhodopsin family)	1.2e-25	83.6
320	7tm_1	7 transmembrane receptor (rhodopsin family)	1.9e-95	305.4
321	7tm_1	7 transmembrane receptor (rhodopsin family)	3.3e-19	63.2
322	TPR	TPR Domain	4.8e-16	66.7
326	Clq	Clq domain	2.7e-31	117.4
330	7tm_1	7 transmembrane receptor (rhodopsin family)	4.3e-15	50.1
333	UbiA	UbiA prenyltransferase family	1.5e-62	221.3
338	7tm_1	7 transmembrane receptor (rhodopsin family)	5.6e-38	122.8
339	7tm_1	7 transmembrane receptor (rhodopsin family)	5.6e-38	122.8
340	COesterase	Carboxylesterase	3.9e-134	459.0
341	7tm_2	7 transmembrane receptor (Secretin family)	2.3e-21	84.4
342	7tm_1	7 transmembrane receptor (rhodopsin family)	3.8e-25	82.1
344	7tm_1	7 transmembrane receptor (rhodopsin family)	1.3e-31	102.6
345	7tm_2	7 transmembrane receptor (Secretin family)	3.3e-73	256.6
346	7tm_2	7 transmembrane receptor (Secretin family)	3.3e-73	256.6
351	ig	Immunoglobulin domain	6.6e-07	27.3
355	LRR	Leucine Rich Repeat	6.1e-29	109.6
357	Reprolysin	Reprolysin (M12B) family zinc metalloprotease	3.7e-93	322.9
358	ig	Immunoglobulin domain	2.7e-08	31.8
359	ig	Immunoglobulin domain	2.7e-08	31.8
362	ig	Immunoglobulin domain	4.1e-08	31.2
365	Folate carrier	Reduced folate carrier	3.5e-145	495.7
368	p450	Cytochrome P450	4.4e-57	203.1
370	gla	Vitamin K-dependent carboxylation/gamma-carboxylglutamic (GLA) domain	6.1e-15	63.1
371	actin	Actin	5.7e-27	89.8
375	TruB_N	TruB family pseudouridylate synthase (N terminal domain)	6.6e-69	242.3
376	TruB_N	TruB family pseudouridylate synthase (N terminal domain)	6.6e-69	242.3
377	abhydrolase	alpha/beta hydrolase fold	0.015	15.7
378	abhydrolase	alpha/beta hydrolase fold	1.1e-10	49.0

Table 4A

SEQ ID NO:	Pfam Model	Description	E-value	Score
382	TTL	Tubulin-tyrosine ligase family	4.1e-122	419.1
383	UQ_con	Ubiquitin-conjugating enzyme	0.0067	-45.5
388	Amino oxidase	Flavin containing amine oxidase	1.3e-17	71.9
389	RUN	RUN domain	8e-51	182.3
390	Rhomboid	Rhomboid family	4.7e-05	30.2
392	Occludin	Occludin/ELL family	1.2e-11	46.2
393	DUF6	Integral membrane protein DUF6	0.037	14.8
395	Patched	Patched family	5.2e-105	362.3
396	zf-C4	Zinc finger, C4 type (two domains)	1.4e-44	152.5
398	Na H Exchanger	Sodium/hydrogen exchanger family	9.9e-103	354.7
402	F-box	F-box domain	0.022	21.4
404	PAP2	PAP2 superfamily	1.4e-30	115.0
406	Patched	Patched family	5.8e-17	-4.9
411	7tm_1	7 transmembrane receptor (rhodopsin family)	5.4e-43	138.7
412	7tm_1	7 transmembrane receptor (rhodopsin family)	2.8e-91	292.1
415	E1-E2 ATPase	E1-E2 ATPase	1.1e-116	387.9
418	HCO3_cotransp	HCO3- transporter family	1.2e-302	1018.9
421	Kelch	Kelch motif	6.5e-40	146.0
422	WD40	WD domain, G-beta repeat	7.5e-16	66.1
423	Beach	Beige/BEACH domain	7.3e-23	86.9
424	bZIP	bZIP transcription factor	0.0074	15.5
430	pkinase	Protein kinase domain	1.8e-36	134.6
432	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	9.4e-06	22.9
434	PMP22 Claudin	PMP-22/EMP/MP20/Claudin family	1.7e-39	144.7
438	MORN	MORN repeat	1.4e-34	128.3
443	PAP2	PAP2 superfamily	2.9e-29	110.7
448	hormone_rec	Ligand-binding domain of nuclear hormone receptor	1e-41	139.0
449	cadherin	Cadherin domain	1.6e-37	138.1
451	zf-CXXC	CXXC zinc finger	2.1e-06	34.7
452	HLH	Helix-loop-helix DNA-binding domain	2.6e-09	44.4
457	ig	Immunoglobulin domain	0.0098	13.9
458	7tm_1	7 transmembrane receptor (rhodopsin family)	1.2e-25	83.6
463	TUDOR	Tudor domain	6.6e-13	56.3
464	Reprolysin	Reprolysin (M12B) family zinc metalloprotease	3.1e-88	306.6
468	HEAT	HEAT repeat	0.0013	25.4
469	DUF6	Integral membrane protein DUF6	1.4e-05	32.0
471	DENN	DENN (AEX-3) domain	7.1e-59	209.0
474	Synaptophysin	Synaptophysin / synaptoporin	4.2e-38	140.0
476	zf-MYND	MYND finger	4.4e-05	29.5
477	7tm_1	7 transmembrane receptor (rhodopsin family)	2.4e-33	108.1
481	HCO3_cotransp	HCO3- transporter family	0	1065.8
482	ank	Ank repeat	1e-19	79.0
485	LRRCT	Leucine rich repeat C-terminal domain	1.1e-08	42.3
486	7tm_1	7 transmembrane receptor (rhodopsin family)	5.3e-42	135.6
490	mito_carr	Mitochondrial carrier protein	5.6e-24	93.1
491	7tm_1	7 transmembrane receptor (rhodopsin family)	3.8e-28	91.6
493	LRR	Leucine Rich Repeat	1.7e-15	64.9
499	Rap_GAP	Rap/ran-GAP	2e-20	81.3
500	fn3	Fibronectin type III domain	1.1e-12	55.6
501	hormone_rec	Ligand-binding domain of nuclear hormone receptor	2e-46	154.4
503	RhoGEF	RhoGEF domain	2.8e-33	124.0
504	fn3	Fibronectin type III domain	1.5e-09	45.1



196

Table 4A

SEQ ID NO:	Pfam Model	Description	E-value	Score
505	7tm_1	7 transmembrane receptor (rhodopsin family)	3.1e-45	145.8
507	trypsin	Trypsin	7e-87	276.1
508	PKD	PKD domain	1.2e-09	45.5
509	Clq	Clq domain	2.7e-31	117.4
513	7tm_1	7 transmembrane receptor (rhodopsin family)	3.3e-12	40.9
516	LRR	Leucine Rich Repeat	7.3e-31	116.0
519	7tm_2	7 transmembrane receptor (Secretin family)	2.3e-21	84.4
521	SNF	Sodium:neurotransmitter symporter family	1.7e-124	427.0
523	SPRY	SPRY domain	9.8e-20	79.0
524	7tm_1	7 transmembrane receptor (rhodopsin family)	5.3e-59	189.6
527	Patched	Patched family	0.00021	-419.9
531	7tm_1	7 transmembrane receptor (rhodopsin family)	3.1e-18	60.1
532	7tm_1	7 transmembrane receptor (rhodopsin family)	1.7e-37	121.3
533	7tm_1	7 transmembrane receptor (rhodopsin family)	6.7e-10	33.6
534	7tm_2	7 transmembrane receptor (Secretin family)	3.3e-73	256.6
535	Rhomboid	Rhomboid family	8.5e-18	72.6
536	Rhomboid	Rhomboid family	8.5e-18	72.6
538	7tm_1	7 transmembrane receptor (rhodopsin family)	4.6e-38	123.1
542	SEA	SEA domain	5.1e-10	46.7
543	SPRY	SPRY domain	2.6e-17	70.9
544	Ribosomal_S26e	Ribosomal protein S26e	2.1e-20	81.2
547	fn3	Fibronectin type III domain	4.1e-102	352.6
548	gla	Vitamin K-dependent carboxylation/gamma-carboxyglutamic (GLA) domain	3e-15	64.1
550	7tm_1	7 transmembrane receptor (rhodopsin family)	4e-43	139.1
551	HCO3_cotransp	HCO3- transporter family	0	1704.8
552	DUF6	Integral membrane protein DUF6	0.069	10.4

197

Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
277	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	1.6e-07	38.5	1	222-263
277	PA	PA domain	1.4e-06	35.3	1	58-144
277	PHD	PHD-finger	0.019	5.9	1	221-266
278	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	1.6e-07	38.5	1	198-239
278	PA	PA domain	0.004	21.3	1	28-120
278	PHD	PHD-finger	0.019	5.9	1	197-242
279	PA	PA domain	1.4e-18	75.2	1	58-162
281	Cornichon	Cornichon protein	4.4e-37	136.6	1	2-113
281	PsbT	Photosystem II reaction centre T protein	3.8	6.4	1	1-24
282	transmembrane 4	Tetraspanin family	1.6e-24	94.9	1	10-166
286	sugar tr	Sugar (and other) transporter	3.9	-186.5	1	19-494
286	Na_sulph_sym p	Sodium:sulfate symporter transmembrane	9	-362.5	1	78-453
287	sushi	Sushi domain (SCR repeat)	1.8e-56	201.1	4	35-94:99-157:162-223:228-283
290	ART	NAD:arginine ADP-ribosyltransferase	1.8e-207	702.6	1	1-326
291	PAP2	PAP2 superfamily	1.3	-21.2	1	88-175
292	UPAR_LY6	u-PAR/Ly-6 domain	0.0034	12.8	1	23-108
292	Keratin B2	Keratin, high sulfur B2 protein	0.48	-63.3	1	7-124
293	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	9.4e-06	32.5	1	7-169
294	MHC_II_alpha	Class II histocompatibility antigen, alp	4.1e-44	160.0	1	29-109
294	ig	Immunoglobulin domain	0.016	21.8	1	125-172
295	Amidase	Amidase	2.1e-65	230.7	1	69-513
296	Na_sulph_sym p	Sodium:sulfate symporter transmembrane	4.1e-71	249.7	1	3-579
296	Na_H_antiporter	Na+/H+ antiporter family	3.3	-108.5	1	241-572
296	Peptidase_C20	Type IV leader peptidase family	6.8	-187.4	1	1-307
296	PHO4	Phosphate transporter family	9	-206.1	1	129-510
298	ABC_membrane	ABC transporter transmembrane region	1.7e-56	201.1	1	188-459
298	ABC_tran	ABC transporter	1.2e-53	191.7	1	469-653
298	APS_kinase	Adenylylsulfate kinase	2.6	-117.0	1	468-587
298	DUF258	Protein of unknown function, DUF258	3.6	-79.4	1	446-596
299	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	0.048	-29.1	1	4-168
300	Mtc	Tricarboxylate carrier	1.2e-67	238.1	1	1-236
301	Mab-21	Mab-21 protein	2.3	-192.1	1	189-524
304	Cornichon	Cornichon protein	3.4e-19	77.2	1	2-98
304	PsbT	Photosystem II reaction centre T protein	3.8	6.4	1	1-24
305	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	1.6	-55.5	1	1-192
306	Acyltransferase	Acyltransferase	4.9e-05	30.2	1	70-229

198  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
308	sugar_tr	Sugar (and other) transporter	0.33	-155.6	1	9-490
308	PUCC	PUCC protein	0.6	-253.1	1	93-486
308	Nucleoside_transport	Nucleoside transporter	2.1	-151.4	1	143-456
308	oxidored_q1	NADH-Ubiquinone/plastoquinone	7	-168.7	1	151-478
309	7tm_1	7 transmembrane receptor (rhodopsin family)	7.1e-05	-4.8	1	41-235
311	Neur_chan_LBD	Neurotransmitter-gated ion-channel lig	1.4e-85	297.7	1	30-236
311	Neur_chan_membrane	Neurotransmitter-gated ion-channel tra	6.5e-38	139.4	1	243-446
312	ig	Immunoglobulin domain	2.1e-17	71.3	3	37-106:138-208:245-300
313	LRR	Leucine Rich Repeat	1.3e-23	91.9	7	66-89:90-113:114-137:138-161:163-186:187-210:211-233
313	ig	Immunoglobulin domain	2.7e-07	37.7	1	314-372
313	fn3	Fibronectin type III domain	2.4e-06	34.5	1	422-502
313	LRRCT	Leucine rich repeat C-terminal domain	5.6e-05	30.0	1	252-297
313	LRRNT	Leucine rich repeat N-terminal domain	3.7	8.7	1	33-64
313	APS_kinase	Adenylylsulfate kinase	5.6	-120.4	1	541-646
314	PSI	Plexin repeat	0.02	20.2	1	303-348
315	PSI	Plexin repeat	0.02	20.2	1	303-348
316	7tm_1	7 transmembrane receptor (rhodopsin family)	4.7e-19	76.7	1	3-245
316	DUF40	Domain of unknown function DUF40	3.1	-127.1	1	2-206
317	Filamin	Filamin/ABP280 repeat	5.5	-34.0	1	100-192
318	Polysacc_synt	Polysaccharide biosynthesis protein	7	-87.4	1	107-368
320	7tm_1	7 transmembrane receptor (rhodopsin family)	1.2e-90	314.5	1	54-335
321	7tm_1	7 transmembrane receptor (rhodopsin family)	2.6e-08	41.0	1	32-309
321	7tm_5	7TM chemoreceptor	8.3	-169.8	1	14-317
322	TPR	TPR Domain	4.3e-16	66.9	3	493-526:527-560:561-594
322	PMT	Dolichyl-phosphate-mannose-protein mannosylt	3.2	-54.0	1	6-245
326	C1q	C1q domain	7.3e-32	119.3	1	117-241
326	Collagen	Collagen triple helix repeat (20 copies)	3.8e-06	33.8	1	50-109
326	Lysis_col	Lysis protein	9.3	-10.9	1	1-36
330	7tm_1	7 transmembrane receptor	0.027	-64.6	1	1-183

199  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		(rhodopsin family)				
331	PKD	PKD domain	1.7e-08	41.7	4	407-495:502-591:596-685:690-782
331	REJ	REJ domain	0.99	-314.6	1	327-806
331	fn3	Fibronectin type III domain	3.7	-2.3	1	408-486
331	Arthro_defensin	Arthropod defensin	4.6	4.0	1	879-907
333	UbiA	UbiA prenyltransferase family	3.2e-56	200.2	1	86-351
338	7tm_1	7 transmembrane receptor (rhodopsin family)	1.1e-34	128.7	1	40-289
338	EII-Sor	PTS system sorbose-specific iic component	9.1	-143.4	1	20-226
339	7tm_1	7 transmembrane receptor (rhodopsin family)	1.1e-34	128.7	1	40-289
339	EII-Sor	PTS system sorbose-specific iic component	9.1	-143.4	1	20-226
340	COesterase	Carboxylesterase	2.3e-133	456.4	1	19-624
341	7tm_2	7 transmembrane receptor	2.3e-21	84.4	1	637-897
341	GPS	Latrophilin/CL-1-like GPS domain	2.7e-13	57.6	1	581-634
341	HRM	Hormone receptor domain	0.0085	15.8	1	298-351
341	Me-amine-deh_L	Methylamine dehydrogenase, L chain	4	-30.1	1	190-321
342	7tm_1	7 transmembrane receptor (rhodopsin family)	3.4e-06	25.9	1	41-225
342	DUF32	Domain of unknown function DUF32	1.9	-145.9	1	37-242
342	DUF40	Domain of unknown function DUF40	9.1	-135.5	1	26-240
344	7tm_1	7 transmembrane receptor (rhodopsin family)	2.2e-28	107.8	1	44-293
344	Abi	CAAX amino terminal protease family	5.4	-25.4	1	101-190
345	7tm_2	7 transmembrane receptor	3.3e-73	256.6	1	396-739
345	GPS	Latrophilin/CL-1-like GPS domain	3.1e-15	64.0	1	345-394
345	metalthio	Metallothionein	1.7	-4.1	1	33-100
345	7tm_5	7TM chemoreceptor	1.7	-157.4	1	392-650
345	CbiM	CbiM	2.1	-83.3	1	497-654
345	DUF26	Domain of unknown function DUF26	2.9	-12.6	1	64-109
345	cytochrome_b_C	Cytochrome b(C-terminal)/b6/petD	4	-28.5	1	369-471
345	TIL	Trypsin Inhibitor like cysteine rich d	9.7	-15.4	1	23-74
346	7tm_2	7 transmembrane receptor	3.3e-73	256.6	1	300-643
346	GPS	Latrophilin/CL-1-like GPS domain	3.1e-15	64.0	1	249-298
346	7tm_5	7TM chemoreceptor	1.7	-157.4	1	296-554
346	CbiM	CbiM	2.1	-83.3	1	401-558
346	cytochrome_b_C	Cytochrome b(C-terminal)/b6/petD	4	-28.5	1	273-375

200  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
351	ig	Immunoglobulin domain	0.00033	27.4	1	72-150
355	LRR	Leucine Rich Repeat	4.6e-29	110.0	7	49-72:73-96:97-120:121-144:146-169:170-193:194-217
355	fn3	Fibronectin type III domain	2.7e-08	41.0	1	387-470
355	ig	Immunoglobulin domain	2.4e-07	37.9	1	278-336
355	LRRCT	Leucine rich repeat C-terminal domain	0.054	17.5	1	218-262
355	LRRNT	Leucine rich repeat N-terminal domain	1	12.9	1	16-47
356	thioredo	Thioredoxin	0.0088	-10.1	1	172-279
357	Reprolysin	Reprolysin (M12B) family zinc metallo	3.6e-93	322.9	1	211-409
357	Pep_M12B_pro pep	Reprolysin family propeptide	7.7e-43	155.7	1	80-196
357	disintegrin	Disintegrin	2.2e-25	97.8	1	426-501
357	Adeno_E3_CR 2	Adenovirus E3 region protein CR2	5.1	-2.5	1	698-738
357	EB	EB module	9.3	-12.3	1	633-682
358	ig	Immunoglobulin domain	6.7e-07	36.4	2	115-168:208-265
359	ig	Immunoglobulin domain	6.7e-07	36.4	2	109-162:202-259
362	ig	Immunoglobulin domain	6.9e-07	36.3	2	47-139:179-274
365	Folate_carrier	Reduced folate carrier	3.8e-145	495.6	1	10-441
365	ion trans	Ion transport protein	8.3	-13.4	1	85-337
365	Nucleoside_transporter	Nucleoside transporter	8.7	-163.1	1	113-367
365	FecCD	FecCD transport family	9.4	-220.8	1	274-457
365	sugar tr	Sugar (and other) transporter	9.7	-198.0	1	11-459
368	p450	Cytochrome P450	4.6e-19	76.8	1	60-379
370	gla	Vitamin K-dependent carboxylation/gamma-carb	3.5e-15	63.9	1	57-98
371	actin	Actin	1.6e-12	55.0	1	8-371
372	DUF140	Domain of unknown function DUF140	5.9	-162.8	1	1-204
375	TruB_N	TruB family pseudouridylate synthase	6.6e-69	242.3	1	107-247
375	PUA	PUA domain	5e-18	73.3	1	339-414
376	TruB_N	TruB family pseudouridylate synthase	6.6e-69	242.3	1	78-218
376	PUA	PUA domain	1.8e-25	98.0	1	266-341
377	abhydrolase	alpha/beta hydrolase fold	0.015	15.7	1	80-270
377	Lipase_3	Lipase (class 3)	0.6	-26.8	1	68-184
377	Thioesterase	Thioesterase domain	1.9	-44.1	1	53-270
378	abhydrolase	alpha/beta hydrolase fold	1.1e-10	49.0	1	80-326
378	Lipase_3	Lipase (class 3)	0.98	-29.1	1	68-198

Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
378	Thioesterase	Thioesterase domain	1.6	-43.0	1	53-297
382	TTL	Tubulin-tyrosine ligase family	1.5e-120	413.9	1	468-764
383	UQ_con	Ubiquitin-conjugating enzyme	4.2e-10	47.0	1	249-412
384	sugar tr	Sugar (and other) transporter	1.2	-171.7	1	54-471
384	voltage CLC	Voltage gated chloride channel	9.2	-243.0	1	92-393
388	Amino_oxidase	Flavin containing amine oxidoreductase	1.9e-69	244.2	1	23-497
389	DENN	DENN (AEX-3) domain	2.1e-87	303.8	1	202-390
389	RUN	RUN domain	8e-51	182.3	1	801-946
389	uDENN	uDENN domain	1.2e-32	121.9	1	4-138
389	dDENN	dDENN domain	3.2e-31	117.1	1	512-588
389	PLAT	PLAT/LH2 domain	7.4e-17	69.4	1	957-1059
390	Rhomboid	Rhomboid family	4.7e-05	30.2	1	59-214
390	UIM	Ubiquitin interaction motif	2.1	14.6	1	268-285
392	Occludin	Occludin/ELL family	1.1e-05	-92.9	1	183-550
392	7tm 5	7TM chemoreceptor	4	-164.0	1	184-451
393	DUF6	Integral membrane protein DUF6	0.042	15.4	1	80-186
393	Nramp	Natural resistance-associated macrophage pro	5.3	-290.4	1	123-381
393	EII-GUT	PTS system enzyme II sorbitol-specific facto	5.8	-135.7	1	192-300
395	Patched	Patched family	3.2e-105	363.0	1	166-965
395	Srg	C.elegans Srg family integral membrane prote	2.7	-213.3	1	214-464
395	UPF0132	Uncharacterised protein family (UPF0132)	4.8	-39.8	1	402-494
395	Sec62	Translocation protein Sec62	5.6	-132.6	1	311-502
396	zf-C4	Zinc finger, C4 type (two domains)	1.8e-42	154.5	1	100-174
396	hormone_rec	Ligand-binding domain of nuclear hormone	7e-17	69.5	1	281-441
398	Na_H_Exchange	Sodium/hydrogen exchanger family	9.9e-103	354.7	1	62-478
398	ABC2_membrane	ABC-2 type transporter	0.92	-112.6	1	254-479
398	GntP_permease	GntP family permease	4.9	-374.7	1	64-366
398	Transp_cyt_pur	Permease for cytosine/purines, uracil	5	-194.9	1	50-427
398	ABC-3	ABC 3 transport family	7.8	-194.6	1	260-469
398	TrkH	Sodium transport protein	7.9	-214.7	1	12-411
398	DUF6	Integral membrane protein DUF6	8	-23.3	1	338-462
398	ER_lumen_receptor	ER lumen protein retaining receptor	8.7	-158.2	1	274-435
399	DUF284	Eukaryotic protein of unknown function, DUF2	1.5e-114	394.0	1	68-309
402	F-box	F-box domain	0.0091	22.6	1	8-55
404	PAP2	PAP2 superfamily	1.4e-30	115.0	1	129-283
406	Patched	Patched family	5.8e-17	-4.9	1	1-756
406	oxidored_q1	NADH-Ubiquinone/plastoquinone (complex I)	0.55	-146.0	1	77-319
406	UPF0118	Domain of unknown function	9.3	-133.5	1	377-719

202  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		DUF20				
411	7tm_1	7 transmembrane receptor (rhodopsin family)	7.1e-38	139.3	1	41-290
411	7tm_5	7TM chemoreceptor	6.7	-168.1	1	16-258
412	7tm_1	7 transmembrane receptor (rhodopsin family)	1.3e-85	297.9	1	43-297
412	7tm_5	7TM chemoreceptor	1.8	-157.8	1	51-305
413	PHD	PHD-finger	0.21	-3.5	1	150-199
413	zf-MIZ	MIZ zinc finger	3.9	-18.2	1	150-200
415	E1-E2 ATPase	E1-E2 ATPase	1.7e-113	390.5	1	223-454
415	Cation_ATPase_C	Cation transporting ATPase, C-terminu	1.7e-69	244.3	1	921-1099
415	Cation_ATPase_N	Cation transporter/ATPase, N-terminus	4.2e-42	153.3	1	121-204
415	Hydrolase	haloacid dehalogenase-like hydrolase	3.7e-15	63.8	1	458-825
415	7tm_5	7TM chemoreceptor	9.4	-170.7	1	170-438
416	MAPEG	MAPEG family	2.1	-21.7	1	98-183
416	Cation_ATPase_C	Cation transporting ATPase, C-terminu	5.6	-47.5	1	81-221
418	HCO3_cotransp	HCO3- transporter family	0	1024.4	1	84-853
418	xan_ur_permease	Permease family	0.9	-176.0	1	375-836
421	Kelch	Kelch motif	3.9e-49	176.7	5	258-308:310-355:357-417:419-471:473-519
421	BTB	BTB/POZ domain	0.88	-10.1	1	2-70
422	WD40	WD domain, G-beta repeat	1.6e-20	81.6	4	16-56:62-98:162-199:313-349
422	aminotran 1 2	Aminotransferase class I and II	0.0091	-46.1	1	391-597
422	Cys_Met_Meta_PP	Cys/Met metabolism PLP-dependent enzy	9.6	-318.8	1	371-600
423	ribonuc_red_sm	Ribonucleotide reductase, small chain	5.6	-142.1	1	989-1265
424	DUF87	Domain of unknown function DUF87	3.9	-134.3	1	48-354
427	DUF6	Integral membrane protein DUF6	3.8	-17.8	1	143-271
427	Frizzled	Frizzled/Smoothed family membrane regio	7.2	-246.3	1	79-280
427	oxidored_q1	NADH-Ubiquinone/plastoquinone (complex I)	9	-170.9	1	70-270
428	DUF6	Integral membrane protein DUF6	3.8	-17.8	1	143-271
428	Frizzled	Frizzled/Smoothed family membrane regio	7.2	-246.3	1	79-280
428	oxidored_q1	NADH-Ubiquinone/plastoquinone	9	-170.9	1	70-270

203  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		(complex I)				
430	pkina	Protein kinase domain	5.6e-33	123.0	1	9-273
432	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.0015	24.7	1	13-59
432	FYVE	FYVE zinc finger	9.5	-26.0	1	10-65
434	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	1.7e-39	144.7	1	89-266
434	Grp1_Fun34_YaaH	GPR1/FUN34/yaaH family	5.9	-120.3	1	71-240
435	DnaJ_CXXCXGXG	DnaJ central domain (4 repeats)	3.5	-46.2	1	37-92
437	AT hook	AT hook motif	3.1	10.6	1	713-725
438	MORN	MORN repeat	1.4e-34	128.3	7	15-37:39-60:61-81:107-129:130-152:288-310:311-333
443	PAP2	PAP2 superfamily	2.9e-29	110.7	1	82-230
448	hormone_rec	Ligand-binding domain of nuclear hormone	3.6e-39	143.6	1	148-329
448	zf-C4	Zinc finger, C4 type (two domains)	3.3e-25	97.2	1	9-66
449	cadherin	Cadherin domain	3.2e-37	137.1	4	15-108:127-227:241-331:342-441
449	SMP-30	Senescence marker protein-30 (SMP-30)	9	-180.9	1	223-467
450	spectrin	Spectrin repeat	0.86	-8.7	1	97-203
451	zf-CXXC	CXXC zinc finger	2.1e-06	34.7	1	131-172
452	HLH	Helix-loop-helix DNA-binding domain	4.4e-09	43.6	1	106-165
453	TP2	Nuclear transition protein 2	8.8	-60.2	1	200-335
458	7tm_1	7 transmembrane receptor (rhodopsin family)	2.1e-05	7.3	1	41-233
463	TUDOR	Tudor domain	6.6e-13	56.3	1	13-134
464	Reprolysin	Reprolysin (M12B) family zinc metallo	3e-88	306.6	1	146-345
464	Pep_M12B_pro pep	Reprolysin family propeptide	1.3e-31	118.4	1	16-134
464	disintegrin	Disintegrin	2.5e-23	90.9	1	362-437
464	EGF	EGF-like domain	0.65	16.5	1	589-616
464	metalthio	Metallothionein	8.7	-12.3	1	362-428
466	SAC3_GANP	SAC3/GANP family	8.8e-77	268.5	1	159-358
468	HEAT	HEAT repeat	0.0012	25.5	1	546-584
469	DUF6	Integral membrane protein DUF6	0.00028	27.7	2	50-179:197-327
469	PhaG_MnhG_YufB	Na <sup>+</sup> /H <sup>+</sup> antiporter subunit	2	-50.3	1	66-168
469	DUF7	Integral membrane protein DUF7	3.9	-34.6	1	227-318



Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
469	Competence	Competence protein	7.5	-104.9	1	93-330
471	DENN	DENN (AEX-3) domain	4.9e-87	302.6	1	57-241
471	dDENN	dDENN domain	1.4e-25	98.4	1	286-353
471	uDENN	uDENN domain	0.0068	-0.5	1	1-50
474	Synaptophysin	Synaptophysin / synaptoporin	4.2e-38	140.0	1	25-241
476	zf-MYND	MYND finger	3e-05	30.9	1	296-335
476	SET	SET domain	2.3	-50.9	1	450-577
476	Antifreeze	Antifreeze-like domain	8.4	-10.3	1	246-295
477	7tm_1	7 transmembrane receptor (rhodopsin family)	2.4e-30	114.2	1	44-293
481	HCO3_cotransp	HCO3- transporter family	0	1072.8	1	108-891
481	xan_ur_permease	Permease family	0.64	-172.1	1	410-874
482	ank	Ankyrin repeat	9.3e-20	79.1	4	172-207;219-251;266-299;345-377
485	LRRCT	Leucine rich repeat C-terminal domain	9.7e-09	42.5	1	9-58
485	GPS	Latrophilin/CL-1-like GPS domain	0.0012	25.4	1	519-571
485	7tm_2	7 transmembrane receptor (Secretin family)	0.0055	-90.7	1	578-784
485	ig	Immunoglobulin domain	0.0078	22.8	1	79-148
485	HRM	Hormone receptor domain	0.069	6.8	1	168-241
486	7tm_1	7 transmembrane receptor	2.9e-38	140.6	1	32-278
486	7tm_5	7TM chemoreceptor	0.23	-141.7	1	55-268
486	V1R	Vomerolateral organ pheromone receptor fami	0.4	-145.6	1	42-291
486	oxidored_q1	NADH-Ubiquinone/plastoquinone (complex I)	4.1	-164.0	1	20-268
486	UPF0032	MitB family UPF0032	7.3	-94.8	1	54-248
490	mito_carr	Mitochondrial carrier protein	6e-24	93.0	2	61-152;155-232
491	7tm_1	7 transmembrane receptor (rhodopsin family)	5.3e-26	99.8	1	41-289
493	LRR	Leucine Rich Repeat	1.2e-15	65.5	5	95-118;119-142;143-166;167-190;191-214
493	LRRNT	Leucine rich repeat N-terminal domain	3e-08	40.9	1	64-93
493	LRRCT	Leucine rich repeat C-terminal domain	7.8e-07	36.1	1	224-277
494	Retrotrans_gag	Retrotransposon gag protein	2	-5.1	1	180-273
495	CDP-OH P transf	CDP-alcohol phosphatidyltransferase	5.8e-08	39.9	1	94-242
495	Cons_hypoth698	Conserved hypothetical protein 698	3	-173.7	1	136-379

205  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
497	oxidored_q1_C	NADH-Ubiquinone oxidoreductase	7.2	-66.0	1	27-276
499	Rap_GAP	Rap/ran-GAP	1.7e-21	84.9	1	1335-1514
500	fn3	Fibronectin type III domain	1.1e-12	55.6	1	47-130
501	hormone_rec	Ligand-binding domain of nuclear hormone	2e-45	164.4	1	364-545
501	zf-C4	Zinc finger, C4 type (two domains)	1.4e-16	68.5	1	269-316
502	7tm_5	7TM chemoreceptor	4.3	-164.6	1	9-304
503	RhoGEF	RhoGEF domain	2.7e-33	124.0	1	320-502
504	fn3	Fibronectin type III domain	1.5e-09	45.1	2	174-267:473-560
505	7tm_1	7 transmembrane receptor (rhodopsin family)	1.7e-41	151.3	1	83-332
505	7tm_5	7TM chemoreceptor	4.5	-165.1	1	89-327
505	DUF40	Domain of unknown function DUF40	4.8	-130.6	1	79-274
506	PFEMP	Plasmodium falciparum erythrocyte membrane p	0.16	-65.7	1	919-1028
507	trypsin	Trypsin	2.6e-79	276.9	1	218-559
507	SRCR	Scavenger receptor cysteine-rich domain	6.2	-22.5	1	120-207
508	PKD	PKD domain	2.6e-09	44.4	1	641-732
508	BNR	BNR/Asp-box repeat	1e-06	35.7	5	54-65:102-113:338-349:415-426:457-468
509	C1q	C1q domain	7.3e-32	119.3	1	211-335
509	Collagen	Collagen triple helix repeat (20 copies)	3.8e-06	33.8	1	144-203
509	Lysis_col	Lysis protein	9.3	-10.9	1	95-130
513	7tm_1	7 transmembrane receptor	1.7e-10	48.3	1	43-294
513	Competence	Competence protein	6.8	-104.0	1	197-459
513	Na_H_antiporter	Na+/H+ antiporter family	8.9	-119.1	1	126-404
514	7tm_5	7TM chemoreceptor	1	-153.5	1	164-454
514	sugar_tr	Sugar (and other) transporter	2.8	-182.4	1	50-547
515	Peptidase_C20	Type IV leader peptidase family	3.3	-182.3	1	99-278
515	MadM	Malonate/sodium symporter MadM subunit	4.7	-20.6	1	209-271
516	LRR	Leucine Rich Repeat	4.8e-31	116.6	8	114-137:138-161:162-184:185-208:209-230:231-254:255-278:279-302
516	LRRNT	Leucine rich repeat N-terminal domain	0.00038	27.2	1	24-55

206  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
516	7tm_1	7 transmembrane receptor	0.0032	-43.2	1	434-683
516	EII-Sor	PTS system sorbose-specific iic compon	5.8	-140.2	1	427-629
516	Cytidyltrans	Phosphatidate cytidyltransferase	7.1	-89.9	1	515-612
516	oxidored_q1	NADH-Ubiquinone/plastoquinone	9.7	-171.5	1	470-680
516	MerC	MerC mercury resistance protein	9.8	-87.5	1	529-627
519	7tm_2	7 transmembrane receptor	2.3e-21	84.4	1	504-764
519	GPS	Latrophilin/CL-1-like GPS domain	2.7e-13	57.6	1	448-501
519	HRM	Hormone receptor domain	0.0085	15.8	1	165-218
519	Me-amine-deh_L	Methylamine dehydrogenase, L chain	4	-30.1	1	57-188
521	SNF	Sodium:neurotransmitter symporter family	4.3e-20	7.1	1	61-289
523	SPRY	SPRY domain	6.1e-20	79.7	1	153-284
524	7tm_1	7 transmembrane receptor (rhodopsin family)	1.6e-52	187.9	1	75-338
524	V1R	Vomer nasal organ pheromone receptor family	7.7	-169.0	1	82-351
525	DUF284	Eukaryotic protein of unknown function, DUF2	2.1e-113	390.1	1	53-350
526	7tm_1	7 transmembrane receptor (rhodopsin family)	0.037	-67.9	1	71-379
527	Patched	Patched family	0.00021	-419.9	1	1-484
528	PSS	Phosphatidyl serine synthase	7.3	-242.7	1	115-277
529	Acyltransferase	Acyltransferase	0.27	-15.8	1	352-517
531	7tm_1	7 transmembrane receptor (rhodopsin family)	0.0063	-49.9	1	96-253
532	7tm_1	7 transmembrane receptor (rhodopsin family)	8.6e-35	129.0	1	62-311
534	7tm_2	7 transmembrane receptor	3.3e-73	256.6	1	179-522
534	GPS	Latrophilin/CL-1-like GPS domain	2.8e-15	64.2	1	128-177
534	7tm_5	7TM chemoreceptor	1.7	-157.4	1	175-433
534	CbiM	CbiM	2.1	-83.3	1	280-437
534	cytochrome_b_C	Cytochrome b(C-terminal)/b6/petD	4	-28.5	1	152-254
535	Rhomboid	Rhomboid family	8.5e-18	72.6	1	647-789
535	Competence	Competence protein	4.4	-100.3	1	640-849
536	Rhomboid	Rhomboid family	8.5e-18	72.6	1	670-812
536	Competence	Competence protein	4.4	-100.3	1	663-872
538	7tm_1	7 transmembrane receptor (rhodopsin family)	6.5e-34	126.1	1	41-290
542	SEA	SEA domain	5.1e-10	46.7	1	472-591
542	EGF	EGF-like domain	0.57	16.7	2	425-462:633-672
542	EB	EB module	4.8	-9.1	1	412-462
542	Bowman-Birk leg	Bowman-Birk serine protease inhibitor	7.2	-18.4	1	628-672
542	Keratin B2	Keratin, high sulfur B2 protein	8.8	-83.0	1	254-385
543	SPRY	SPRY domain	7.8e-17	69.4	1	347-468

207  
Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
543	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	3.1e-11	50.7	1	16-56
543	zf-B_box	B-box zinc finger	5.7e-05	29.9	1	92-133
544	Ribosomal_S26e	Ribosomal protein S26e	2.1e-20	81.2	1	1-110
544	rnaseA	Pancreatic ribonuclease	1.3e-07	32.0	1	106-232
545	Patched	Patched family	0.33	-525.2	1	37-846
545	oxidored_q3	NADH-ubiquinone/plastoquinone oxidoreduct	4.3	-79.9	1	201-368
545	oxidored_q1	NADH-Ubiquinone/plastoquinone (complex I)	9.7	-171.5	1	663-851
545	Keratin_B2	Keratin, high sulfur B2 protein	10	-83.9	1	11-141
546	7tm_1	7 transmembrane receptor (rhodopsin family)	0.028	-65.2	1	47-249
547	fn3	Fibronectin type III domain	4.1e-102	352.6	6	947-1034:1046-1138:1150-1239:1251-1337:1444-1527:1541-1623
547	ig	Immunoglobulin domain	1.8e-87	304.0	9	199-260:300-356:389-448:482-547:579-637:670-731:764-829:863-929:1364-1425
548	gla	Vitamin K-dependent carboxylation/gamma-carb	3.7e-15	63.8	1	24-65
550	7tm_1	7 transmembrane receptor (rhodopsin family)	1.1e-39	145.3	1	41-290
550	DUF40	Domain of unknown function DUF40	2	-123.7	1	39-229
551	HCO3_cotransp	HCO3- transporter family	0	1723.0	1	146-959
551	xan_ur_permease	Permease family	3.3	-190.7	1	477-941
551	Plant_vir_prot	Plant virus coat protein	9.3	-51.7	1	772-865
551	DENN	DENN (AEX-3) domain	9.5	-71.3	1	593-719
552	DUF6	Integral membrane protein DUF6	0.092	9.6	1	68-174
552	DUF250	Domain of unknown function, DUF250	2.8	-98.0	1	180-351
552	oxidored_q3	NADH-ubiquinone/plastoquinone	5.9	-82.1	1	81-236

208

Table 4B

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		oxidoreduct				
552	7tm 5	7TM chemoreceptor	9.2	-170.6	1	54-338

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
277	1g25	A	218	270	1.2e-17	0.00	-0.05		CDK-ACTIVATING KINASE ASSEMBLY FACTOR MAT1; CHAIN: A;	METAL BINDING PROTEIN RING FINGER PROTEIN MAT1; RING FINGER (C3HC4)
278	1g25	A	194	246	1.2e-17	0.00	-0.05		CDK-ACTIVATING KINASE ASSEMBLY FACTOR MAT1; CHAIN: A;	METAL BINDING PROTEIN RING FINGER PROTEIN MAT1; RING FINGER (C3HC4)
294	1a1a	A	2	187	6.8e-77	-0.48	0.01		B*3501; CHAIN: A, B; PEPTIDE VPLRPMTY; CHAIN: C;	COMPLEX (ANTIGEN/PEPTIDE) B35; MAJOR HISTOCOMPATIBILITY ANTIGEN, MHC, HLA, HLA-B3501, HIV, 2 NEF, COMPLEX (ANTIGEN/PEPTIDE)
294	1a6a	B	28	187	6.8e-56	-0.23	0.52		HLA-DR3; CHAIN: A, B; CLIP; CHAIN: C;	COMPLEX (TRANSMEMBRANE/GLYCOPROT EIN) MHC GLYCOPROTEIN, COMPLEX (TRANSMEMBRANE/GLYCOPROT EIN)
294	1a6a	B	7	188	6.8e-56			65.70	HLA-DR3; CHAIN: A, B; CLIP; CHAIN: C;	COMPLEX (TRANSMEMBRANE/GLYCOPROT EIN) MHC GLYCOPROTEIN, COMPLEX (TRANSMEMBRANE/GLYCOPROT EIN)
294	1agd	A	2	187	3.4e-78	-0.52	0.00		B*0801; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; HIV-1 GAG PEPTIDE (GGKKYKL - INDEX PEPTIDE); CHAIN: C;	HISTOCOMPATIBILITY COMPLEX B8; B2M; PEPTIDE HLA B8, HIV, MHC CLASS I, HISTOCOMPATIBILITY COMPLEX
294	1agd	B	19	186	8.5e-56			63.94	HLA-DRI CLASS II	COMPLEX (MHC)

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									HISTOCOMPATIBILITY PROTEIN; CHAIN: A, B, D, E, G, H, J, K; HLA-A2; CHAIN: C, F, L, L;	PROTEIN/ANTIGEN) DRA, DRB1 01010; COMPLEX (MHC PROTEIN/ANTIGEN), HISTOCOMPATIBILITY ANTIGEN
294	1a9d	B	28	187	8.5e-56	-0.35	0.62		HLA-DRI CLASS II HISTOCOMPATIBILITY PROTEIN; CHAIN: A, B, D, E, G, H, J, K; HLA-A2; CHAIN: C, F, L, L;	COMPLEX (MHC PROTEIN/ANTIGEN) DRA, DRB1 01010; COMPLEX (MHC PROTEIN/ANTIGEN), HISTOCOMPATIBILITY ANTIGEN
294	1bx2	A	27	188	8.4e-64			251.62	HLA-DR2; CHAIN: A, D; HLA-DR2; CHAIN: B, E; HLA-DR2; CHAIN: C, F;	IMMUNE SYSTEM HLA-DR2, MYELIN BASIC PROTEIN, MULTIPLE SCLEROSIS, 2 AUTOIMMUNITY, IMMUNE SYSTEM
294	1bx2	A	27	188	8.4e-64	0.21	1.00		HLA-DR2; CHAIN: A, D; HLA-DR2; CHAIN: B, E; HLA-DR2; CHAIN: C, F;	IMMUNE SYSTEM HLA-DR2, MYELIN BASIC PROTEIN, MULTIPLE SCLEROSIS, 2 AUTOIMMUNITY, IMMUNE SYSTEM
294	1bx2	B	27	190	6.8e-57			61.42	HLA-DR2; CHAIN: A, D; HLA-DR2; CHAIN: B, E; HLA-DR2; CHAIN: C, F;	IMMUNE SYSTEM HLA-DR2, MYELIN BASIC PROTEIN, MULTIPLE SCLEROSIS, 2 AUTOIMMUNITY, IMMUNE SYSTEM
294	1bx2	B	35	190	6.8e-57	-0.27	0.59		HLA-DR2; CHAIN: A, D; HLA-DR2; CHAIN: B, E; HLA-DR2; CHAIN: C, F;	IMMUNE SYSTEM HLA-DR2, MYELIN BASIC PROTEIN, MULTIPLE SCLEROSIS, 2 AUTOIMMUNITY, IMMUNE SYSTEM
294	1duz	A	2	186	3.4e-73	-0.46	0.00		HLA-A*0201; CHAIN: A, D; BETA-2 MICROGLOBULIN;	IMMUNE SYSTEM IMMUNOGLOBULIN FOLD

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CHAIN: B, E; HTLV-1 OCTAMERIC TAX PEPTIDE; CHAIN: C, F;	
294	1f3j	B	27	188	3.4e-54	-0.43	0.24		H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN; CHAIN: A, D; MHC CLASS II NOD; CHAIN: B, E; LYSOZYME C; CHAIN: P, Q;	IMMUNE SYSTEM HISTOCOMPATIBILITY ANTIGEN, MHC, PEPTIDE COMPLEX
294	1fv1	B	28	187	1e-55	-0.37	0.17		MAJOR HISTOCOMPATIBILITY COMPLEX ALPHA CHAIN; CHAIN: A, D; MAJOR HISTOCOMPATIBILITY COMPLEX BETA CHAIN; CHAIN: B, E; MYELIN BASIC PROTEIN; CHAIN: C, F;	IMMUNE SYSTEM MHC CLASS II DR2A
294	1iak	B	28	186	8.5e-53			55.66	MHC CLASS II I-AK; CHAIN: A, B, P; HEN EGGWHITE LYSOZYME PEPTIDE	HISTOCOMPATIBILITY ANTIGEN I-AK HISTOCOMPATIBILITY ANTIGEN, MHC, PEPTIDE COMPLEX
294	1iao	B	18	186	5.1e-55			57.77	MHC CLASS II I-AD; CHAIN: A, B;	MHC II MHC II, CLASS II MHC, I-A, OVALBUMIN PEPTIDE
294	1iao	B	24	185	5.1e-55	-0.44	0.57		MHC CLASS II I-AD; CHAIN: A, B;	MHC II MHC II, CLASS II MHC, I-A, OVALBUMIN PEPTIDE
294	1ica	B	1	185	6.8e-53			66.30	MHC CLASS II I-EK; CHAIN: A, B, C, D;	HISTOCOMPATIBILITY ANTIGEN HISTOCOMPATIBILITY ANTIGEN
294	1ica	B	22	185	6.8e-53	-0.37	0.69		MHC CLASS II I-EK; CHAIN: A, B, C, D;	HISTOCOMPATIBILITY ANTIGEN HISTOCOMPATIBILITY ANTIGEN
294	1ieb	B	1	185	6.8e-53			70.67	MHC CLASS II I-EK; CHAIN: A, B, C, D;	HISTOCOMPATIBILITY ANTIGEN HISTOCOMPATIBILITY ANTIGEN



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
294	1ieb	B	22	185	6.8e-53	-0.55	0.65		MHC CLASS II I-EK; CHAIN: A, B, C, D;	HISTOCOMPATIBILITY ANTIGEN HISTOCOMPATIBILITY ANTIGEN
294	1qo3	A	2	185	8.5e-77	-0.67	0.00		MHC CLASS I H-2DD HEAVY CHAIN; CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; HIV ENVELOPE GLYCOPROTEIN 120 PEPTIDE; CHAIN: P; LY49A; CHAIN: C, D;	HISTOCOMPATIBILITY ANTIGEN CLASS II H-2 CLASS I HISTOCOMPATIBILITY ANTIGEN, B2M; NK-CELL SURFACE GLYCOPROTEIN YE1/48, NK CELL, INHIBITORY RECEPTOR, MHC-I, C-TYPE LECTIN-LIKE, 2 HISTOCOMPATIBILITY, B2M, LY49, LY-49
294	1qgd	A	3	185	1.7e-75	-0.43	0.03		HISTOCOMPATIBILITY LEUKOCYTE ANTIGEN (HLA)-CW4 CHAIN: A; BETA-2-MICROGLOBULIN; CHAIN: B; HLA-CW4 SPECIFIC PEPTIDE; CHAIN: C;	IMMUNE SYSTEM IMMUNOGLOBULIN (IG)-LIKE DOMAIN, ALPHA HELIX, BETA SHEET, 2 IMMUNE SYSTEM
294	2iad	B	15	186	3.4e-55			59.90	MHC CLASS II I-AD; CHAIN: A, B;	MHC II MHC II, CLASS II MHC I-AD
294	2iad	B	27	185	3.4e-55	-0.10	0.22		MHC CLASS II I-AD; CHAIN: A, B;	MHC II MHC II, CLASS II MHC I-AD
313	1ady	A	65	282	1.1e-15	0.26	0.66		RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE), COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPITOPE MAPPING, LEUCINE-RICH 3 REPEATS
313	1a9n	A	144	295	8.4e-11	0.08	-0.13		U2 RNA HAIRPIN IV; CHAIN:	COMPLEX (NUCLEAR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	PROTEIN/RNA COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1a9n	A	65	153	8.4e-09	0.10	0.17		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1a9n	A	71	217	2.4e-22	0.38	0.40		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1a9n	A	96	226	7.2e-19	0.09	-0.05		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1a9n	C	144	295	3.6e-11	0.16	-0.08		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1a9n	C	65	162	3.6e-09	0.14	0.19		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1a9n	C	71	222	4.8e-22	0.48	0.49		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
313	1b1h	A	297	387	8.4e-15	0.42	0.72		HEMOLIN; CHAIN: A, B;	INSECT IMMUNITY INSECT IMMUNITY, LPS-BINDING, HOMOPHILIC ADHESION
313	1cvs	D	311	402	9.6e-15	-0.07	0.33		FIBROBLAST GROWTH	GROWTH FACTOR/GROWTH

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
313	1d0b	A	161	313	5.1e-24	-0.04	0.09		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
313	1d0b	A	40	193	3.4e-18	0.18	0.10		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
313	1d0b	A	59	227	3.6e-18	0.16	0.40		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
313	1d0b	A	85	265	1.7e-19	-0.05	0.11		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
313	1dce	A	40	149	6.8e-08	-0.13	0.76		RAB GERANYLGERANYLTRANSFERASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFERASE BETA SUBUNIT; CHAIN: B, D;	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFERASE, 2.0 A 2 RESOLUTION, N-FORMYL METHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
313	1ds9	A	138	227	1.2e-11	-0.75	0.09		OUTER ARM DYNEIN; CHAIN: A;	CONTRACTILE PROTEIN LEUCINE-RICH REPEAT, BETA-BETA-ALPHA CYLINDER, DYNEIN, 2 CHLAMYDOMONAS, FLAGELLA
313	1ds9	A	69	216	7.2e-12	-0.26	0.22		OUTER ARM DYNEIN; CHAIN: A;	CONTRACTILE PROTEIN LEUCINE-RICH REPEAT, BETA-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
313	1ds9	A	87	203	2.4e-18	0.24	1.00		OUTER ARM DYNEIN; CHAIN: A;	BETA-ALPHA CYLINDER, DYNEIN, 2 CHLAMYDOMONAS, FLAGELLA CONTRACTILE PROTEIN
313	1ern	A	428	500	0.00036	-0.31	0.49		ERYTHROPOIETIN RECEPTOR; CHAIN: A, B;	ERYTHROPOIETIN RECEPTOR, SIGNAL TRANSDUCTION, CYTOKINE 2 RECEPTOR CLASS 1
313	1ev2	E	311	409	3.6e-15	-0.08	0.18		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
313	1ev2	G	311	402	4.8e-15	0.00	0.82		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
313	1ev1	C	261	388	2.4e-16	-0.32	0.05		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
313	1fhg	A	292	387	1.2e-18	0.30	0.80		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
313	1fna		428	517	2.4e-10	-0.01	0.88		CELL ADHESION PROTEIN FIBRONECTIN CELL- ADHESION MODULE TYPE III-10 IFNA 3	
313	1fo1	A	42	101	6.8e-05	-0.72	0.00		NUCLEAR RNA EXPORT FACTOR 1; CHAIN: A, B;	RNA BINDING PROTEIN TAP (NFX1); RIBONUCLEOPROTEIN (RNP,RBD OR RRM) AND LEUCINE-RICH-REPEAT 2 (LRR)
313	1fqv	A	46	283	1e-10	0.06	-0.17		SKP2; CHAIN: A, C, E, G, I, K, M, O; SKP1; CHAIN: B, D, F, H, J, L, N, P;	LIGASE CYCLIN A/CDK2- ASSOCIATED PROTEIN P45; CYCLIN A/CDK2-ASSOCIATED PROTEIN P19; SKP1, SKP2, F-BOX, LRR, LEUCINE-RICH REPEAT, SCF, UBIQUITIN, 2 E3, UBIQUITIN PROTEIN LIGASE
313	1fs2	A	59	230	3.6e-15	0.04	-0.11		SKP2; CHAIN: A, C; SKP1; CHAIN: B, D;	LIGASE CYCLIN A/CDK2- ASSOCIATED P45; CYCLIN A/CDK2-ASSOCIATED P19; SKP1, SKP2, F-BOX, LRRS, LEUCINE- RICH REPEATS, SCF, 2 UBIQUITIN, E3, UBIQUITIN PROTEIN LIGASE
313	1mfu		424	507	2.4e-07	-0.33	0.09		FIBRONECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN, RGD, EXTRACELLULAR MATRIX, 2 HEPARIN-BINDING, GLYCOPROTEIN
313	1mm		311	387	2.4e-17	0.44	0.69		MUSCLE PROTEIN TITIN MODULE M5 (CONNECTIN) ITNM 3 (NMR, MINIMIZED AVERAGE STRUCTURE) ITNM 4 ITNM 58	

217

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
313	1ttf		426	517	7.2e-10	-0.24	0.84		GLYCOPROTEIN FIBRONECTIN (TENTH TYPE III MODULE) (NMR, 36 STRUCTURES) 1TTF 3	
313	2bth		66	230	9.6e-22	0.12	0.30		RIBONUCLEASE INHIBITOR; CHAIN: NULL;	ACETYLATION RNASE INHIBITOR, RIBONUCLEASE/ANGIOGENIN INHIBITOR ACETYLATION, LEUCINE-RICH REPEATS
313	2fcb	A	232	387	1.2e-12	0.12	0.82		FC GAMMA RIIB; CHAIN: A;	IMMUNE SYSTEM CD32: RECEPTOR, FC, CD32, IMMUNE SYSTEM
313	2ncm		298	387	3.6e-15	0.05	0.45		NEURAL CELL ADHESION MOLECULE; CHAIN: NULL;	CELL ADHESION NCAM DOMAIN 1; CELL ADHESION, GLYCOPROTEIN, HEPARIN-BINDING, GPI-ANCHOR, 2 NEURAL ADHESION MOLECULE, IMMUNOGLOBULIN FOLD, SIGNAL
322	1a17		471	580	7.2e-12	0.09	0.16		SERINE/THREONINE PROTEIN PHOSPHATASE 5; CHAIN: NULL;	HYDROLASE TETRAPEPTIDE, TRP; HYDROLASE, PHOSPHATASE, PROTEIN-PROTEIN INTERACTIONS, TPR, 2 SUPER-HELIX, X-RAY STRUCTURE
322	1a17		490	620	1.7e-25	-0.03	0.15		SERINE/THREONINE PROTEIN PHOSPHATASE 5; CHAIN: NULL;	HYDROLASE TETRAPEPTIDE, TRP; HYDROLASE, PHOSPHATASE, PROTEIN-PROTEIN INTERACTIONS, TPR, 2 SUPER-HELIX, X-RAY STRUCTURE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
322	1e96	B	367	525	3.4e-10	0.08	-0.18		RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1; CHAIN: A; NEUTROPHIL CYTOSOL FACTOR 2 (NCF-2) CHAIN: B;	SIGNALLING COMPLEX RAC1; P67PHOX; SIGNALLING COMPLEX, GTPASE, NADPH OXIDASE, PROTEIN-PROTEIN 2 COMPLEX, TPR MOTIF
322	1e96	B	461	616	1.2e-12	-0.05	0.21		RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1; CHAIN: A; NEUTROPHIL CYTOSOL FACTOR 2 (NCF-2) CHAIN: B;	SIGNALLING COMPLEX RAC1; P67PHOX; SIGNALLING COMPLEX, GTPASE, NADPH OXIDASE, PROTEIN-PROTEIN 2 COMPLEX, TPR MOTIF
322	1e96	B	492	635	1e-13	0.03	0.33		RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1; CHAIN: A; NEUTROPHIL CYTOSOL FACTOR 2 (NCF-2) CHAIN: B;	SIGNALLING COMPLEX RAC1; P67PHOX; SIGNALLING COMPLEX, GTPASE, NADPH OXIDASE, PROTEIN-PROTEIN 2 COMPLEX, TPR MOTIF
322	1e1r	A	467	554	3.4e-12	-0.23	0.13		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE MEEVD; CHAIN: B;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSP90, 2 PROTEIN BINDING
322	1e1r	A	497	588	3.4e-17	-0.02	0.66		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE MEEVD; CHAIN: B;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSP90, 2 PROTEIN BINDING
322	1e1r	A	533	629	6.8e-14	0.11	0.15		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE MEEVD; CHAIN: B;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSP90, 2 PROTEIN BINDING
322	1e1w	A	425	533	1e-13	0.00	-0.15		TPR1-DOMAIN OF HOP; CHAIN: A, B; HSC70- PEPTIDE; CHAIN: C, D;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSC70, 2 HSP70, PROTEIN BINDING

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
322	1elw	A	464	562	1.7e-11	0.15	0.30		TPR1-DOMAIN OF HOP; CHAIN: A, B; HSC70- PEPTIDE; CHAIN: C, D;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSC70, 2 HSP70, PROTEIN BINDING
322	1elw	A	494	607	5.1e-22	0.23	0.90		TPR1-DOMAIN OF HOP; CHAIN: A, B; HSC70- PEPTIDE; CHAIN: C, D;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSC70, 2 HSP70, PROTEIN BINDING
322	1elw	A	530	634	6.8e-14	-0.02	0.27		TPR1-DOMAIN OF HOP; CHAIN: A, B; HSC70- PEPTIDE; CHAIN: C, D;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSC70, 2 HSP70, PROTEIN BINDING
322	1fch	A	405	635	1.4e-35	-0.18	0.00		PEROXISOMAL TARGETING SIGNAL 1 RECEPTOR; CHAIN: A, B; PTS1- CONTAINING PEPTIDE; CHAIN: C, D;	SIGNALING PROTEIN PEROXISOMAL RECEPTOR 1, PTS1- BP, PEROXIN-5, PTS1 PROTEIN- PEPTIDE COMPLEX, TETRAPEPTIDE REPEAT, TPR, 2 HELICAL REPEAT
322	1qge	A	493	623	8.4e-07	-0.11	0.17		VESICULAR TRANSPORT PROTEIN SEC17; CHAIN: A;	PROTEIN TRANSPORT HELIX- TURN-HELIX TPR-LIKE REPEAT, PROTEIN TRANSPORT
341	1dan	L	35	151	1.7e-09	0.04	-0.20		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE- ARG- CHLOROMETHYLKETONE (DEFRMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO- FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
341	1dva	L	114	206	3.4e-10	0.06	-0.17		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE



220

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	COMPLEX
341	1emn		113	184	3.4e-08	0.07	-0.15		FIBRILLIN; CHAIN: NULL;	MATRIX PROTEIN EXTRACELLULAR MATRIX, CALCIUM-BINDING, GLYCOPROTEIN, 2 REPEAT, SIGNAL, MULTIGENE FAMILY, DISEASE MUTATION, 3 EGF-LIKE DOMAIN, HUMAN FIBRILLIN-1 FRAGMENT, MATRIX PROTEIN
341	1xka	L	114	191	6.8e-09	0.18	-0.05		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
345	1apo		70	108	6.8e-09	0.07	0.13		COAGULATION FACTOR EGF-LIKE MODULE OF BLOOD COAGULATION FACTOR X (N-TERMINAL, I APO 3 APO FORM) (NMR, 13 STRUCTURES) I APO 4	
345	1aut	L	30	102	3.4e-12	0.36	0.04		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAL; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE, PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
345	1dan	L	27	119	1.7e-11	-0.10	0.37		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLOROMETHYLKETONE (DFRCMK) WITH CHAIN: C;	COMPLEX (BLOOD COAGULATION/INHIBITOR)
345	1dva	L	27	119	1.7e-11	-0.15	0.27		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
345	1dva	L	70	161	1.5e-16	-0.05	0.04		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
345	1emm		27	103	5.1e-14	0.29	0.16		FIBRILLIN; CHAIN: NULL;	MATRIX PROTEIN EXTRACELLULAR MATRIX, CALCIUM-BINDING, GLYCOPROTEIN, 2 REPEAT, SIGNAL, MULTIGENE FAMILY, DISEASE MUTATION, 3 EGF-LIKE DOMAIN, HUMAN FIBRILLIN-1 FRAGMENT, MATRIX PROTEIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
345	1f7e	A	70	112	1.7e-08	0.10	0.49		BLOOD COAGULATION FACTOR VII; CHAIN: A;	BLOOD CLOTTING FACTOR VII, BLOOD COAGULATION, EGF-LIKE DOMAIN, BLOOD 2 CLOTTING
345	1fak	L	27	119	1.7e-11	-0.02	0.28		BLOOD COAGULATION FACTOR VIIA; CHAIN: L; BLOOD COAGULATION FACTOR VIIA; CHAIN: H; SOLUBLE TISSUE FACTOR; CHAIN: T; 5L15; CHAIN: I;	BLOOD CLOTTING COMPLEX(SERINE PROTEASE/COFACTOR/LIGAND), BLOOD COAGULATION, 2 SERINE PROTEASE, COMPLEX, CO-FACTOR, RECEPTOR ENZYME, 3 INHIBITOR, GLA, EGF, COMPLEX (SERINE 4 PROTEASE/COFACTOR/LIGAND), BLOOD CLOTTING
345	1pfx	L	27	119	1.2e-10	0.07	0.11		FACTOR IXA; CHAIN: C, L; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
345	1pfx	L	70	158	3.4e-12	0.27	0.10		FACTOR IXA; CHAIN: C, L; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
345	1qfk	L	29	119	5.1e-11	-0.22	0.31		COAGULATION FACTOR VIIA (LIGHT CHAIN);	SERINE PROTEASE FVILA; FVILA; BLOOD COAGULATION, SERINE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	PROTEASE
345	1xka	L	29	119	3.4e-11	0.13	0.58		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
358	1adq	L	3	178	5.1e-15	0.25	-0.17		IGG4 REA; CHAIN: A; RF-AN IGM/LAMBDA; CHAIN: H, L;	COMPLEX (IMMUNOGLOBULIN/AUTOANTIGEN) COMPLEX (IMMUNOGLOBULIN/AUTOANTIGEN), RHEUMATOID FACTOR 2 AUTO-ANTIBODY COMPLEX
358	1b4j	H	69	288	0.00034			50.30	ANTIBODY; CHAIN: L, H;	ANTIBODY ENGINEERING ANTIBODY ENGINEERING, HUMANIZED AND CHIMERIC ANTIBODIES, 2 FAB, X-RAY STRUCTURES, GAMMA-INTERFERON
358	1b1h	A	14	270	6.8e-31	0.03	-0.17		HEMOLIN; CHAIN: A, B;	INSECT IMMUNITY INSECT IMMUNITY, LPS-BINDING, HOMOPHILIC ADHESION
358	1cvs	C	112	281	5.1e-38	0.34	0.17		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
358	1cvs	D	112	281	3.4e-40	0.17	0.05		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	FACTOR/GROWTH FACTOR RECEPTOR
358	1cvs	D	12	184	3.4e-25	0.04	-0.14		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
358	1epf	A	105	266	5.1e-20	0.25	0.36		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM, NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
358	1ev2	E	112	281	5.1e-35	0.15	0.15		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2, FGFR2, IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
358	1ev2	G	112	281	1.7e-37	0.24	0.58		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2, FGFR2, IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD

225

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQ FOLD score	Compound	PDB annotation
358	1evr	C	112	281	1e-39	0.29	0.13		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF1, FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREEFOLD FOLD
358	1f2q	A	3	193	1.7e-34	-0.05	0.25		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A;	IMMUNE SYSTEM FC-EPSILON RI-ALPHA; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN
358	1f6a	A	3	192	1.7e-34	0.09	0.24		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A; IG EPSILON CHAIN C REGION; CHAIN: B, D;	IMMUNE SYSTEM HIGH AFFINITY IGE-FC RECEPTOR, FC(EPSILON) IGE-FC; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN, IGE ANTIBODY, IGE-FC
358	1f6a	A	98	287	3.4e-25	0.17	0.75		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A; IG EPSILON CHAIN C REGION; CHAIN: B, D;	IMMUNE SYSTEM HIGH AFFINITY IGE-FC RECEPTOR, FC(EPSILON) IGE-FC; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN, IGE ANTIBODY, IGE-FC
358	1fcg	A	105	284	1.7e-26	0.35	0.63		FC RECEPTOR FC(GAMMA)R1A; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32
358	1fcg	A	2	190	5.1e-36	0.24	0.59		FC RECEPTOR FC(GAMMA)R1A; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32

226

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
358	1fhg	A	99	181	5.1e-12	0.46	-0.15		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
358	1fml	A	2	189	1.7e-33	0.06	0.06		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A;	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
358	1h1l	A	107	266	3.4e-10	0.02	-0.06		IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) 1H1L 3	
358	1ifh	L	107	266	3.4e-10	0.11	-0.02		IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) COMPLEX WITH PEPTIDE OF 1IFH 3 INFLUENZA HEMAGGLUTININ HA1 (STRAIN X47) (RESIDUES 101-107) 1IFH 4	
358	1ibb	B	25	270	1.7e-18	-0.26	0.05		INTERLEUKIN-1 BETA; CHAIN: A; TYPE 1 INTERLEUKIN-1 RECEPTOR; CHAIN: B;	COMPLEX (IMMUNOGLOBULIN/RECEPTOR) IMMUNOGLOBULIN FOLD, TRANSMEMBRANE, GLYCOPROTEIN, RECEPTOR, 2 SIGNAL, COMPLEX (IMMUNOGLOBULIN/RECEPTOR)
358	1koa		100	194	1.7e-10	0.50	-0.07		TWITCHIN; CHAIN: NULL;	KINASE KINASE, TWITCHIN, INTRASTERIC REGULATION
358	1mco	H	1	379	8.5e-10			57.76	IMMUNOGLOBULIN IMMUNOGLOBULIN G1 (IGG1) (MCG) WITH A HINGE DELETION 1MCO 3	
358	1nct		104	181	8.5e-13	0.21	-0.12		TITIN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN,

227

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										NEXTM5; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN FOLD, ALTERNATIVE SPLICING, SIGNAL, 3 MUSCLE PROTEIN
358	1nkr		100	284	3.4e-38	0.41	0.98		P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
358	1nkr		103	284	3.4e-38			72.82	P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
358	1tnm		104	181	8.5e-13	0.04	-0.12		MUSCLE PROTEIN TTIN MODULE M5 (CONNECTIN) ITNM 3 (NMR, MINIMIZED AVERAGE STRUCTURE) ITNM 4 ITNM 58	
358	2dli	A	99	283	5.1e-38	0.26	0.66		MHC CLASS I NK CELL RECEPTOR PRECURSOR; CHAIN: A;	IMMUNE SYSTEM P58 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN
358	2fcb	A	105	286	6.8e-26	0.33	0.82		FC GAMMA RIIB; CHAIN: A;	IMMUNE SYSTEM CD32: RECEPTOR, FC, CD32, IMMUNE SYSTEM



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
358	2fc6	A	2	189	1.5e-36	0.14	0.69		FC GAMMA RIIB; CHAIN: A;	IMMUNE SYSTEM CD32; RECEPTOR, FC, CD32, IMMUNE SYSTEM
359	1aif	L	98	260	5.1e-11	0.23	0.05		ANTI-IDIOTYPIC FAB 409.5.3 (IGG2A) FAB; CHAIN: A, B, L, H	IMMUNOGLOBULIN IMMUNOGLOBULIN, C REGION, V REGION
359	1epf	A	99	260	3.4e-21	0.06	0.04		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
359	1ev2	E	95	275	1.7e-37	-0.12	0.07		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2; FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE 1- SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
359	1ev2	G	95	275	3.4e-40	0.15	0.10		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2; FGFR2; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE 1- SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
359	1ev1	C	94	275	6.8e-42	-0.02	0.16		FIBROBLAST GROWTH FACTOR 1; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF1; FGFR1; IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE 1- SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
359	1f2q	A	6	187	1.7e-37	0.17	0.31		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN;	IMMUNE SYSTEM FC-EPSILON RI- ALPHA; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									A;	RECEPTOR, IGE-BINDING 2 PROTEIN
359	1f2q	A	95	282	5.1e-23	0.28	0.27		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A;	IMMUNE SYSTEM FC-EPSILON RI-ALPHA; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN
359	1f6a	A	3	186	1.2e-37	0.09	0.09		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A; IG EPSILON CHAIN C REGION; CHAIN: B, D;	IMMUNE SYSTEM HIGH AFFINITY IGE-FC RECEPTOR, FC(EPSILON) IGE-FC; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN, IGE ANTIBODY, IGE-FC
359	1f6a	A	95	281	5.1e-23	0.09	0.54		HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR CHAIN: A; IG EPSILON CHAIN C REGION; CHAIN: B, D;	IMMUNE SYSTEM HIGH AFFINITY IGE-FC RECEPTOR, FC(EPSILON) IGE-FC; IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IGE-BINDING 2 PROTEIN, IGE ANTIBODY, IGE-FC
359	1f62	D	93	293	1.7e-06			50.05	IMMUNOGLOBULIN IMMUNOGLOBULIN FC AND FRAGMENT B OF PROTEIN A COMPLEX 1FC2 4	
359	1f6g	A	4	184	5.1e-38	0.21	0.25		FC RECEPTOR FC(GAMMA)RIIA; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32
359	1f6g	A	99	278	1.7e-23	0.06	0.36		FC RECEPTOR FC(GAMMA)RIIA; CHAIN: A;	IMMUNE SYSTEM, MEMBRANE PROTEIN CD32; FC RECEPTOR, IMMUNOGLOBULIN, LEUKOCYTE, CD32

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
359	1fhg	A	93	175	1.2e-12	0.46	-0.15		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
359	1fhl	A	106	280	5.1e-21	0.20	0.25		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A:	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
359	1fhl	A	5	183	3.4e-34	0.00	-0.01		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A:	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
359	1fhl	A	97	280	1.1e-22	0.20	0.62		LOW AFFINITY IMMUNOGLOBULIN GAMMA FC REGION CHAIN: A:	IMMUNE SYSTEM RECEPTOR BETA SANDWICH, IMMUNOGLOBULIN-LIKE, RECEPTOR
359	1fth	L	101	260	3.4e-11	0.05	-0.08		IMMUNOGLOBULIN IGG2A FAB FRAGMENT (FAB 17/9) COMPLEX WITH PEPTIDE OF 11FH 3 INFLUENZA HEMAGGLUTININ HA1 (STRAIN X47) (RESIDUES 101-107) 11FH 4	
359	1mco	H	1	373	3.4e-10			54.90	IMMUNOGLOBULIN IMMUNOGLOBULIN G1 (IGG1) (MCG) WITH A HINGE DELETION IMCO 3	
359	1nct		98	175	5.1e-13	0.21	-0.12		TITIN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN, NEXTMS; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
359	1nkr		95	278	1.7e-39	0.26	0.78		P58-CL42 KIR; CHAIN: NULL;	FOLD, ALTERNATIVE SPLICING, SIGNAL, 3 MUSCLE PROTEIN INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR, INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
359	1nkr		97	278	1.7e-39			73.34	P58-CL42 KIR; CHAIN: NULL;	INHIBITORY RECEPTOR KILLER CELL INHIBITORY RECEPTOR; INHIBITORY RECEPTOR, NATURAL KILLER CELLS, IMMUNOLOGICAL 2 RECEPTORS, IMMUNOGLOBULIN FOLD
359	1mm		98	175	5.1e-13	0.04	-0.12		MUSCLE PROTEIN TTIN MODULE M5 (CONNECTIN) 1TNM 3 (NMR, MINIMIZED AVERAGE STRUCTURE) 1TNM 4 1TNM 58	
359	1vca	A	10	186	3.4e-11	0.13	-0.20		HUMAN VASCULAR CELL ADHESION MOLECULE-1; 1VCA 4 CHAIN: A, B; 1VCA 5	CELL ADHESION PROTEIN VCAM-D1, 2; 1VCA 6 IMMUNOGLOBULIN SUPERFAMILY, INTEGRIN-BINDING 1VCA 15
359	1yuh	H	126	345	0.0012			50.43	FAB FRAGMENT; CHAIN: NULL;	IMMUNOGLOBULIN ANTI-NITROPHENOL, LAMBDA LIGHT CHAIN, IMMUNOGLOBULIN
359	2di	A	93	277	1e-37	0.18	0.72		MHC CLASS I NK CELL RECEPTOR PRECURSOR; CHAIN: A;	IMMUNE SYSTEM P58 NATURAL KILLER CELL RECEPTOR; KIR, NATURAL KILLER RECEPTOR, INHIBITORY RECEPTOR, 2 IMMUNOGLOBULIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
359	2fcb	A	4	185	3.4c-38	0.19	0.22		FC GAMMA RIIB; CHAIN: A;	IMMUNE SYSTEM CD32; RECEPTOR, FC, CD32, IMMUNE SYSTEM
359	2fcb	A	99	280	8.5c-24	0.35	0.37		FC GAMMA RIIB; CHAIN: A;	IMMUNE SYSTEM CD32; RECEPTOR, FC, CD32, IMMUNE SYSTEM
362	1bwm	A	47	278	8.8c-10	0.20	-0.12		ALPHA-BETA T CELL RECEPTOR (TCR) (D10); CHAIN: A;	IMMUNE SYSTEM IMMUNOGLOBULIN, IMMUNORECEPTOR, IMMUNE SYSTEM
362	1d9k	B	45	137	8.8c-05	0.24	0.24		T-CELL RECEPTOR D10 (ALPHA CHAIN); CHAIN: A, E; T-CELL RECEPTOR D10 (BETA CHAIN); CHAIN: B, F; MHC I-AK A CHAIN (ALPHA CHAIN); CHAIN: C, G; MHC I-AK B CHAIN (BETA CHAIN); CHAIN: D, H; CONALBUMIN PEPTIDE; CHAIN: P, Q;	IMMUNE SYSTEM MHC I-AK; MHC I-AK; T-CELL RECEPTOR, MHC CLASS II, D10, I-AK
362	1fyt	D	174	300	1.5c-05	0.13	0.05		HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR CHAIN: A; HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR-1 CHAIN: B; HEMAAGLUTININ HA1 PEPTIDE CHAIN; CHAIN: C; T-CELL RECEPTOR ALPHA CHAIN; CHAIN: D; T-CELL RECEPTOR BETA CHAIN;	IMMUNE SYSTEM HLA-DRI, DRA, HLA-DRI, DRB1 0101; TCR HA1.7 ALPHA CHAIN; TCR HA1.7 BETA CHAIN; PROTEIN-PROTEIN COMPLEX, IMMUNOGLOBULIN FOLD

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
362	1nfd	B	38	165	6.6e-06	0.18	0.01		CHAIN: E; N15 ALPHA-BETA T-CELL RECEPTOR; CHAIN: A, B, C, D; H57 FAB; CHAIN: E, F, G, H	COMPLEX (IMMUNORECEPTOR/IMMUNOGL OBULIN) COMPLEX (IMMUNORECEPTOR/IMMUNOGL OBULIN)
362	1wio	A	171	305	6.6e-09	0.41	0.09		T-CELL SURFACE GLYCOPROTEIN CD4; CHAIN: A, B;	GLYCOPROTEIN CD4; IMMUNOGLOBULIN FOLD, TRANSMEMBRANE. GLYCOPROTEIN, T-CELL, 2 MHC LIPOPROTEIN, POLYMORPHISM
371	1dga	A	12	371	0	0.42	1.00		ACTIN; CHAIN: A; GELSOLIN; CHAIN: G;	CONTRACTILE PROTEIN ACTIN, GELSOLIN, CYTOSKELETON ORGANIZATION, ACTIN-2 ASSOCIATED PROTEIN
371	1esv	A	14	371	0	0.60	1.00		GELSOLIN; CHAIN: S; ALPHA ACTIN; CHAIN: A	CONTRACTILE PROTEIN LATRUNCULIN A, GELSOLIN, ACTIN, DEPOLYMERIZATION, 2 SEQUESTRATION
371	1yag	A	12	370	0			148.05	ACTIN; CHAIN: A; GELSOLIN; CHAIN: G;	CONTRACTILE PROTEIN ACTIN- DEPOLYMERIZING FACTOR (ADF); COMPLEX, ACTIN, GELSOLIN, CONTRACTILE PROTEIN
371	1yag	A	12	372	0	0.72	1.00		ACTIN; CHAIN: A; GELSOLIN; CHAIN: G;	CONTRACTILE PROTEIN ACTIN- DEPOLYMERIZING FACTOR (ADF); COMPLEX, ACTIN, GELSOLIN, CONTRACTILE PROTEIN
371	2btf	A	10	371	0	0.48	1.00		ACETYLATION AND ACTIN- BINDING BETA-ACTIN- PROFILIN COMPLEX 2BTF 3	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
371	2btf	A	10	372	0			148.88	ACETYLATION AND ACTIN- BINDING BETA-ACTIN- PROFILIN COMPLEX 2BTF 3	
377	1a88	A	30	266	3.6e-26			69.79	CHLOROPEROXIDASE L; CHAIN: A, B, C;	HALOPEROXIDASE BROMOPEROXIDASE L, HALOPEROXIDASE L; HALOPEROXIDASE OXIDOREDUCTASE
377	1a88	A	32	265	3.6e-26	0.43	0.87		CHLOROPEROXIDASE L; CHAIN: A, B, C;	HALOPEROXIDASE BROMOPEROXIDASE L, HALOPEROXIDASE L; HALOPEROXIDASE OXIDOREDUCTASE
377	1a8q		30	260	1.1e-10	0.11	0.62		BROMOPEROXIDASE A1; CHAIN: NULL;	HALOPEROXIDASE CHLOROPEROXIDASE A1, HALOPEROXIDASE A1; HALOPEROXIDASE, OXIDOREDUCTASE
377	1a8q		30	272	1.1e-10			58.83	BROMOPEROXIDASE A1; CHAIN: NULL;	HALOPEROXIDASE CHLOROPEROXIDASE A1, HALOPEROXIDASE A1; HALOPEROXIDASE, OXIDOREDUCTASE
377	1a8s		30	266	2.4e-25			65.01	CHLOROPEROXIDASE F; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE F; HALOPEROXIDASE, OXIDOREDUCTASE, PROPIONATE COMPLEX
377	1a8s		33	265	2.4e-25	0.23	1.00		CHLOROPEROXIDASE F; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE F;

235

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
377	1a2w	A	21	164	8.4e-18	0.16	0.68		PROLINE IMINOPEPTIDASE; CHAIN: A, B;	HALOPEROXIDASE, OXIDOREDUCTASE, PROPIONATE COMPLEX
377	1b6g		12	265	7.2e-24	0.34	1.00		HALOALKANE DEHALOGENASE; CHAIN: NULL;	HYDROLASE HYDROLASE, HALOALKANE DEHALOGENASE, ALPHA/BETA-HYDROLASE
377	1b7t		30	268	2.4e-23			51.04	BROMOPEROXIDASE A2; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE A2, CHLOROPEROXIDASE A2; HALOPEROXIDASE, OXIDOREDUCTASE, PEROXIDASE, ALPHA/BETA 2 HYDROLASE FOLD, MUTANT M99T
377	1b7t		31	260	2.4e-23	0.22	0.71		BROMOPEROXIDASE A2; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE A2, CHLOROPEROXIDASE A2; HALOPEROXIDASE, OXIDOREDUCTASE, PEROXIDASE, ALPHA/BETA 2 HYDROLASE FOLD, MUTANT M99T
377	1c4x	A	25	265	1.1e-27			52.47	2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE CHAIN: A;	HYDROLASE BPHD; HYDROLASE, PCB DEGRADATION
377	1c4x	A	31	265	1.1e-27	0.47	0.95		2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE CHAIN: A;	HYDROLASE BPHD; HYDROLASE, PCB DEGRADATION



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
377	1din		70	254	0.00048	-0.34	0.52		DIENELACTONE HYDROLASE; CHAIN: NULL;	HYDROLYTIC ENZYME DLH; DIENELACTONE HYDROLASE, AROMATIC HYDROCARBON CATABOLISM, 2 SERINE ESTERASE, CARBOXYMETHYLENEBUTENOLI DASE, 3 HYDROLYTIC ENZYME
377	1ehy	A	31	257	7.2e-15	0.05	0.69		SOLUBLE EPOXIDE HYDROLASE; CHAIN: A, B, C, D;	HYDROLASE HYDROLASE, ALPHA/BETA HYDROLASE FOLD, EPOXIDE DEGRADATION, 2 EPICHLOROHYDRIN
377	1ek1	A	17	167	2.4e-20	0.29	0.70		EPOXIDE HYDROLASE; CHAIN: A, B;	HYDROLASE HOMODIMER, ALPHA/BETA HYDROLASE FOLD, DISUBSTITUTED UREA 2 INHIBITOR
377	1ek1	B	17	167	3.6e-20	0.40	0.65		EPOXIDE HYDROLASE; CHAIN: A, B;	HYDROLASE HOMODIMER, ALPHA/BETA HYDROLASE FOLD, DISUBSTITUTED UREA 2 INHIBITOR
377	1evq	A	3	275	4.8e-11	0.38	0.86		SERINE HYDROLASE; CHAIN: A;	HYDROLASE ALPHA/BETA HYDROLASE FOLD
377	1ex9	A	92	157	1.2e-06	-0.03	0.48		LACTONIZING LIPASE; CHAIN: A;	HYDROLASE TRIACYL- GLYCEROL LIPASE; LIPASE, ALPHA-BETA HYDROLASE FOLD, PSEUDOMONAS, PHOSPHONATE 2 INHIBITOR
377	1hlg	A	20	159	0.0048	0.46	0.83		LIPASE, GASTRIC; CHAIN: A, B;	HYDROLASE LIPASE
377	1qj4	A	53	266	7.2e-30	0.20	0.89		HYDROXYNITRILE LYASE; CHAIN: A;	LYASE OXYNITRILE LYASE; OXYNITRILE, CYANOGENESIS, CYANHYDRIN FORMATION,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										LYASE
378	1a88	A	30	322	1.1e-29			84.31	CHLOROPEROXIDASE L; CHAIN: A, B, C;	HALOPEROXIDASE BROMOPEROXIDASE L, HALOPEROXIDASE L; HALOPEROXIDASE, OXIDOREDUCTASE
378	1a88	A	32	321	1.1e-29	0.11	0.99		CHLOROPEROXIDASE L; CHAIN: A, B, C;	HALOPEROXIDASE BROMOPEROXIDASE L, HALOPEROXIDASE L; HALOPEROXIDASE, OXIDOREDUCTASE
378	1a8q		30	328	9.6e-12			81.71	BROMOPEROXIDASE A1; CHAIN: NULL;	HALOPEROXIDASE CHLOROPEROXIDASE A1, HALOPEROXIDASE A1; HALOPEROXIDASE, OXIDOREDUCTASE
378	1a8s		30	322	7.2e-29			82.20	CHLOROPEROXIDASE F; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE F; HALOPEROXIDASE, OXIDOREDUCTASE, PROPIONATE COMPLEX
378	1a8s		33	321	7.2e-29	0.14	0.96		CHLOROPEROXIDASE F; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE F; HALOPEROXIDASE, OXIDOREDUCTASE, PROPIONATE COMPLEX
378	1a8u	A	47	178	0.00024	0.32	0.60		CARBOXYLESTERASE; CHAIN: A, B;	HYDROLASE HYDROLASE
378	1a2w	A	16	336	8.4e-25			66.76	PROLINE IMINOPEPTIDASE; CHAIN: A, B;	AMINOPEPTIDASE AMINOPEPTIDASE, PROLINE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
378	1a2w	A	21	312	8.4e-25	-0.03	0.34		PROLINE IMINOPEPTIDASE; CHAIN: A, B;	IMINOPEPTIDASE, SERINE PROTEASE, 2 XANTHOMONAS CAMPESTRIS
378	1b6g		12	321	6e-27	0.28	0.82		HALOALKANE DEHALOGENASE; CHAIN: NULL;	HYDROLASE HYDROLASE, HALOALKANE DEHALOGENASE, ALPHA/BETA-HYDROLASE
378	1b6g		5	303	6e-27			70.09	HALOALKANE DEHALOGENASE; CHAIN: NULL;	HYDROLASE HYDROLASE, HALOALKANE DEHALOGENASE, ALPHA/BETA-HYDROLASE
378	1b7t		30	324	6e-26			69.36	BROMOPEROXIDASE A2; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE A2, CHLOROPEROXIDASE A2; HALOPEROXIDASE, OXIDOREDUCTASE, PEROXIDASE, ALPHA/BETA 2 HYDROLASE FOLD, MUTANT M99T
378	1b7t		31	316	6e-26	0.23	0.72		BROMOPEROXIDASE A2; CHAIN: NULL;	HALOPEROXIDASE HALOPEROXIDASE A2, CHLOROPEROXIDASE A2; HALOPEROXIDASE, OXIDOREDUCTASE, PEROXIDASE, ALPHA/BETA 2 HYDROLASE FOLD, MUTANT M99T
378	1c4x	A	25	321	1.2e-30			75.67	2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE CHAIN: A;	HYDROLASE BPHD; HYDROLASE, PCB DEGRADATION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
378	1c4x	A	31	321	1.2e-30	0.23	0.98		2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE CHAIN: A;	HYDROLASE BPHD; HYDROLASE, PCB DEGRADATION
378	1ehy	A	17	294	2.4e-17			51.96	SOLUBLE EPOXIDE HYDROLASE; CHAIN: A, B, C, D;	HYDROLASE HYDROLASE, ALPHA/BETA HYDROLASE FOLD, EPOXIDE DEGRADATION, 2 EPICHLOROHYDRIN
378	1ehy	A	31	313	2.4e-17	-0.01	0.37		SOLUBLE EPOXIDE HYDROLASE; CHAIN: A, B, C, D;	HYDROLASE HYDROLASE, ALPHA/BETA HYDROLASE FOLD, EPOXIDE DEGRADATION, 2 EPICHLOROHYDRIN
378	1ek1	A	17	316	3.6e-29	0.25	0.76		EPOXIDE HYDROLASE; CHAIN: A, B;	HYDROLASE HOMODIMER, ALPHA/BETA HYDROLASE FOLD, DISUBSTITUTED UREA 2 INHIBITOR
378	1ek1	B	17	321	8.4e-30	0.20	0.92		EPOXIDE HYDROLASE; CHAIN: A, B;	HYDROLASE HOMODIMER, ALPHA/BETA HYDROLASE FOLD, DISUBSTITUTED UREA 2 INHIBITOR
378	1evq	A	3	192	2.4e-07	0.12	0.86		SERINE HYDROLASE; CHAIN: A;	HYDROLASE ALPHA/BETA HYDROLASE FOLD
378	1ex9	A	92	174	9.6e-07	-0.14	0.09		LACTONIZING LIPASE; CHAIN: A;	HYDROLASE TRIACYL-GLYCEROL LIPASE; LIPASE, ALPHA-BETA HYDROLASE FOLD, PSEUDOMONAS, PHOSPHONATE 2 INHIBITOR
378	1fg2	A	47	157	4.8e-06	0.14	0.42		ACYL PROTEIN THIOESTERASE 1; CHAIN: A, B;	HYDROLASE ALPHA/BETA HYDROLASE, SERINE HYDROLASE, SAD, ANOMALOUS 2 DIFFRACTION
378	1hlg	A	20	221	0.00084	0.29	0.49		LIPASE, GASTRIC; CHAIN:	HYDROLASE LIPASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
378	1qj4	A	51	305	6e-33			57.61	A, B; HYDROXYNITRILE LYASE; CHAIN: A;	LYASE OXYNITRILE LYASE; OXYNITRILASE, CYANOGENESIS, CYANHYDRIN FORMATION, LYASE
378	1qj4	A	53	322	6e-33	0.25	0.63		HYDROXYNITRILE LYASE; CHAIN: A;	LYASE OXYNITRILE LYASE; OXYNITRILASE, CYANOGENESIS, CYANHYDRIN FORMATION, LYASE
383	1ayz	A	249	415	1.7e-48	0.32	0.80		UBIQUITIN-CONJUGATING ENZYME RAD6; CHAIN: A, B, C;	UBIQUITIN CONJUGATION UBC2; UBIQUITIN CONJUGATION, UBIQUITIN-CONJUGATING ENZYME
383	1ayz	A	249	419	1.7e-48			56.83	UBIQUITIN-CONJUGATING ENZYME RAD6; CHAIN: A, B, C;	UBIQUITIN CONJUGATION UBC2; UBIQUITIN CONJUGATION, UBIQUITIN-CONJUGATING ENZYME
383	1c4z	D	253	412	1.7e-40	0.10	0.07		UBIQUITIN-PROTEIN LIGASE E3A; CHAIN: A, B, C; UBIQUITIN CONJUGATING ENZYME E2; CHAIN: D;	LIGASE E6AP, UBC7; BILOBAL STRUCTURE, ELONGATED SHAPE, E3 UBIQUITIN LIGASE, E2 2 UBIQUITIN CONJUGATING ENZYME
383	1qcq	A	250	406	6.8e-54			57.98	UBIQUITIN CONJUGATING ENZYME; CHAIN: A;	LIGASE UBIQUITIN, UBIQUITIN- CONJUGATING ENZYME, YEAST
383	1qcq	A	251	413	6.8e-54	0.37	0.96		UBIQUITIN CONJUGATING ENZYME; CHAIN: A;	LIGASE UBIQUITIN, UBIQUITIN- CONJUGATING ENZYME, YEAST
383	1u9a	A	249	413	3.4e-43	0.30	1.00		UBC9; CHAIN: NULL;	UBIQUITIN-CONJUGATING ENZYME UBIQUITIN- CONJUGATING ENZYME; UBIQUITIN-CONJUGATING

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PME score	SEQ FOLD score	Compound	PDB annotation
										ENZYME, UBIQUITIN-DIRECTED 2 PROTEOLYSIS, CELL CYCLE CONTROL, LIGASE
383	2aak		249	412	1.4e-48	0.56	1.00		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC1; UBIQUITIN CONJUGATION, LIGASE
383	2aak		249	415	1.4e-48			62.30	UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC1; UBIQUITIN CONJUGATION, LIGASE
383	2e2c		245	419	1.5e-43	0.25	0.84		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBIQUITIN CONJUGATION, UBIQUITIN CARRIER PROTEIN, THIOESTER 2 BOND, LIGASE
383	2e2c		247	422	1.5e-43			57.90	UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBIQUITIN CONJUGATION, UBIQUITIN CARRIER PROTEIN, THIOESTER 2 BOND, LIGASE
383	2uc2		249	412	1.7e-41	-0.00	0.43		UBIQUITIN CONJUGATING ENZYME; CHAIN: NULL;	UBIQUITIN CONJUGATION UBC7; UBIQUITIN CONJUGATION, LIGASE, YEAST
388	1b37	A	32	498	1.7e-52	0.57	1.00		POLYAMINE OXIDASE; CHAIN: A, B, C;	OXIDOREDUCTASE FLAVIN-DEPENDENT AMINE OXIDASE, OXIDOREDUCTASE
388	1f8s	A	32	496	3.4e-51	0.18	0.88		L-AMINO ACID OXIDASE; CHAIN: A, B, C, D, E, F, G, H;	OXIDOREDUCTASE FLAVOENZYME, OXIDASE, ENANTIOERIC SPECIFICITY, O-2 AMINO BENZOATE, ACTIVE SITE FUNNEL, HELICAL DOMAIN, FAD-3 BINDING DOMAIN
388	1qo8	A	16	52	0.0044	-0.29	0.11		FLAVOCYTOCHROME C3	OXIDOREDUCTASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									FUMARATE REDUCTASE; CHAIN: A, D;	OXIDOREDUCTASE
389	1lox		956	1098	3.4c-36	0.25	-0.14		15-LIPOXYGENASE; CHAIN: NULL;	OXIDOREDUCTASE 15LOX; OXIDOREDUCTASE, 15LO DEPOT2
402	1a4y	A	109	458	5.1e-15	-0.23	0.22		RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE), COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPTOPE MAPPING, LEUCINE-RICH 3 REPEATS
402	1a4y	A	27	486	5.1e-15			68.37	RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE), COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPTOPE MAPPING, LEUCINE-RICH 3 REPEATS
402	1fo1	B	214	309	1.7e-09	-0.59	0.00		NUCLEAR RNA EXPORT FACTOR 1; CHAIN: A, B;	RNA BINDING PROTEIN TAP (NFX1); RIBONUCLEOPROTEIN (RNP, RBD OR RRM) AND LEUCINE-RICH-REPEAT 2 (LRR)
402	1fs1	A	10	50	1.2e-12	-0.48	0.78		CYCLIN A/CDK2-ASSOCIATED P19; CHAIN: A, C; CYCLIN A/CDK2-ASSOCIATED P45; CHAIN: B, D;	LIGASE SKP2 F-BOX; SKP1; SKP1, SKP2, F-BOX, LRR, LEUCINE-RICH REPEAT, SCF, UBIQUITIN, 2 E3, UBIQUITIN PROTEIN LIGASE

243

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
402	1fs2	A	9	291	1.5e-40	-0.31	0.24		SKP2; CHAIN: A, C; SKP1; CHAIN: B, D;	LIGASE CYCLIN A/CDK2-ASSOCIATED P45; CYCLIN A/CDK2-ASSOCIATED P19; SKP1, SKP2, F-BOX, LRRS, LEUCINE-RICH REPEATS, SCF, 2 UBIQUITIN, E3, UBIQUITIN PROTEIN LIGASE
402	2bnh		111	458	5.1e-17	-0.33	0.00		RIBONUCLEASE INHIBITOR; CHAIN: NULL;	ACETYLATION RNASE INHIBITOR, RIBONUCLEASE/ANGIOGENIN INHIBITOR ACETYLATION, LEUCINE-RICH REPEATS
410	1cy8		324	479	4e-08	0.15	-0.19		GLYCOSYLTRANSFERASE CYCLODEXTRIN GLUCANOTRANSFERASE (E.C.2.4.1.19) (CGTASE) 1CYG 3	
410	1eci	B	220	250	0.01	-0.47	0.16		ECTATOMIN; 1ECI 5 CHAIN: A, B 1ECI 6	TOXIN PORE-FORMING TOXINS, ANT VENOMS 1ECI 11
413	1chc		149	218	1.7e-12	0.12	-0.12		VIRUS EQUINE HERPES VIRUS-1 (CHC4, OR RING DOMAIN) 1CHC 3 (NMR, 1 STRUCTURE) 1CHC 4	
413	1md		147	208	8.5e-08	-0.04	0.21		RAG1; CHAIN: NULL;	DNA-BINDING PROTEIN V(D)J RECOMBINATION ACTIVATING PROTEIN 1; RAG1, V(D)J RECOMBINATION, ANTIBODY, MAD, RING FINGER, 2 ZINC BINUCLEAR CLUSTER, ZINC FINGER, DNA-BINDING PROTEIN



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
415	1eul	A	131	1102	0	0.29	1.00		CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC CHAIN: A;	HYDROLASE SERCA1; ION PUMP, CALCIUM, MEMBRANE PROTEIN, P-TYPE ATPASE, ACTIVE 2 TRANSPORT
418	1bzk	A	361	401	6e-16	-0.88	0.05		BAND 3 ANION TRANSPORT PROTEIN; CHAIN: A;	TRANSPORT PROTEIN HUMAN ERYTHROCYTE ANION TRANSPORTER, TRANSMEMBRANE, 2 SYNTHETIC PEPTIDE, NMR
419	1aof	A	101	199	4e-06	0.04	0.40		GLUTATHIONE S-TRANSFERASE; CHAIN: A, B;	TRANSFERASE GST, GLUTATHIONE TRANSFERASE; TRANSFERASE, GLUTATHIONE CONUGATTION, DETOXIFICATION, DETOXIFICATION ENZYME GST, CGSTM1-1; DETOXIFICATION ENZYME, GLUTATHIONE S-TRANSFERASE, S-HEXYL 2 GLUTATHIONE
419	1gsu	A	105	199	0.0001	0.08	0.46		CLASS-MU GLUTATHIONE S-TRANSFERASE; CHAIN: A, B;	
419	1gta		121	173	0.00012	-0.54	0.28		GLUTATHIONE TRANSFERASE GLUTATHIONE S-TRANSFERASE (E.C.2.5.1.18) (26 KDA) 1GTA 3	
419	1hna		102	166	0.00016	-0.16	0.31		TRANSFERASE(GLUTATHIONE) GLUTATHIONE S-TRANSFERASE (HUMAN, CLASS MU) (GSTM2-2) 1HNA 3 FORM A (E.C.2.5.1.18) MUTANT WITH TRP 214	

245

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
419	1hna		137	195	1.4e-05	-0.25	0.17		REPLACED BY PHE 1HNA 4 (W214F) 1HNA 5	
									TRANSFERASE(GLUTATHIONE) GLUTATHIONE S-TRANSFERASE (HUMAN, CLASS MU) (GSTM2-2) 1HNA 3 FORM A (E.C.2.5.1.18) MUTANT WITH TRP 214 REPLACED BY PHE 1HNA 4 (W214F) 1HNA 5	
419	1pmt		104	199	4e-05	0.43	0.35		GLUTATHIONE TRANSFERASE; CHAIN: NULL;	TRANSFERASE PMGST, GST B1-1; TRANSFERASE, GLUTATHIONE-CONJUGATING, A PUTATIVE 2 OXIDOREDUCTASE
419	3gtu	B	102	173	0.0002	-0.05	0.10		GLUTATHIONE S-TRANSFERASE; CHAIN: A, B, C, D;	TRANSFERASE TRANSFERASE, GLUTATHIONE, CONJUGATION, DETOXIFICATION, 2 CYTOSOLIC, HETERODIMER
419	4gtu	A	141	195	6.8e-05	-0.03	0.19		GLUTATHIONE S-TRANSFERASE; CHAIN: A, B, C, D, E, F, G, H;	TRANSFERASE TRANSFERASE, GLUTATHIONE, CONJUGATION, DETOXIFICATION, 2 CYTOSOLIC, HOMODIMER
419	6gsv	A	121	199	0.0002	0.09	0.18		MU CLASS GLUTATHIONE S-TRANSFERASE OF ISOENZYMES CHAIN: A, B;	GLUTATHIONE TRANSFERASE RAT GST; GLUTATHIONE TRANSFERASE, ISOENZYME 3-3, T13S MUTANT
420	1aw9		80	325	1.7e-44	0.17	0.62		GLUTATHIONE S-TRANSFERASE III; CHAIN: NULL;	TRANSFERASE TRANSFERASE, HERBICIDE DETOXIFICATION
420	1axd	A	80	317	1.5e-37	0.14	0.23		GLUTATHIONE S-	COMPLEX

246

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									TRANSFERASE I; CHAIN: A, B; LACTOYLGLUTATHIONE; CHAIN: C, D	(TRANSFERASE/LIGAND) COMPLEX (TRANSFERASE/LIGAND), TRANSFERASE, HERBICIDE 2 DETOXIFICATION HEADER
420	1b48	A	71	325	5.1e-31	0.22	-0.08		GLUTATHIONE S-TRANSFERASE; CHAIN: A, B;	TRANSFERASE GST, MGST44-4; CRYSTAL STRUCTURE, GLUTATHIONE S-TRANSFERASE, GST, SUBUNIT 2 COOPERATIVITY
420	1b8x	A	122	372	3.4e-21			57.80	AML-1B; CHAIN: A;	SIGNAL PROTEIN NUCLEAR MATRIX TARGETING SIGNAL PROTEIN
420	1bg5		122	358	3.4e-21			51.42	FUSION PROTEIN OF ALPHA-NA,K-ATPASE WITH CHAIN: NULL;	ANKYRIN BINDING MAB; ANKYRIN BINDING, ATPASE, GLUTATHIONE S-TRANSFERASE, CARRIER 2 CRYSTALLIZATION, ION TRANSPORT
420	1eem	A	81	320	1.4e-33	0.14	-0.12		GLUTATHIONE S-TRANSFERASE; CHAIN: A;	TRANSFERASE GST, GLUTATHIONE CONJUGATING, PUTATIVE OXIDOREDUCTASE
420	1gne		122	336	3.4e-21			55.36	GLUTATHIONE TRANSFERASE GLUTATHIONE S-TRANSFERASE (E.C.2.5.1.18) FUSED WITH A IGNE 3 CONSERVED NEUTRALIZING EPTOPE ON GP41 OF HUMAN IGNE 4 IMMUNODEFICIENCY VIRUS TYPE 1, COMPLEXED WITH GLUTATHIONE IGNE	

247

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
420	1gne		231	331	0.0004	0.25	0.54	5	GLUTATHIONE TRANSFERASE S-GLUTATHIONE S-TRANSFERASE (E.C.2.5.1.18) FUSED WITH A 1GNE 3 CONSERVED NEUTRALIZING EPITOPE ON GP41 OF HUMAN 1GNE 4 IMMUNODEFICIENCY VIRUS TYPE 1, COMPLEXED WITH GLUTATHIONE 1GNE 5	
420	1gnw	A	81	315	1.4e-33	0.28	0.41		GLUTATHIONE S-TRANSFERASE; CHAIN: A, B;	TRANSFERASE TRANSFERASE, HERBICIDE DETOXIFICATION
420	1gse	A	71	335	6.8e-36	0.04	-0.14		GLUTATHIONE TRANSFERASE; 1GSE 6 CHAIN: A, B; 1GSE 7	TRANSFERASE (GLUTATHIONE) A1-1 1GSE 19
420	1gta		122	326	3.4e-21			51.70	GLUTATHIONE TRANSFERASE S-TRANSFERASE (E.C.2.5.1.18) (26 KDA) 1GTA 3	
421	1buo	A	1	66	5.1e-09	-0.01	0.10		PROMYELOCYTIC LEUKEMIA ZINC FINGER PROTEIN PLZF; CHAIN: A;	GENE REGULATION POZ DOMAIN; PROTEIN-PROTEIN INTERACTION DOMAIN, TRANSCRIPTIONAL 2 REPRESSOR, ZINC-FINGER PROTEIN, X-RAY CRYSTALLOGRAPHY, 3 PROTEIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
421	1gof		227	544	3.4e-19	0.10	0.95		OXIDOREDUCTASE(OXYGE N(A)) GALACTOSE OXIDASE (E.C.1.1.3.9) (PH 4.5) 1GOF 3	STRUCTURE, PROMYELOCYTIC LEUKEMIA, GENE REGULATION
422	1bjw	A	409	597	8e-09	0.34	0.82		ASPARTATE AMINOTRANSFERASE; CHAIN: A, B;	AMINOTRANSFERASE AMINOTRANSFERASE, PYRIDOXAL ENZYME
422	1bs0	A	348	559	1.7e-62	-0.05	1.00		8-AMINO-7-OXONANOATE SYNTHASE; CHAIN: A;	TRANSFERASE AONS, 8-AMINO-7-KETOPELARGONATE SYNTHASE; PLP-DEPENDENT ACYL-COA SYNTHASE, BIOTIN BIOSYNTHESIS, 8-2 AMINO-7-OXONANOATE SYNTHASE, 8-AMINO-7-KETOPELARGONATE 3 SYNTHASE, TRANSFERASE
422	1bs0	A	406	596	1.8e-38	0.80	1.00		8-AMINO-7-OXONANOATE SYNTHASE; CHAIN: A;	TRANSFERASE AONS, 8-AMINO-7-KETOPELARGONATE SYNTHASE; PLP-DEPENDENT ACYL-COA SYNTHASE, BIOTIN BIOSYNTHESIS, 8-2 AMINO-7-OXONANOATE SYNTHASE, 8-AMINO-7-KETOPELARGONATE 3 SYNTHASE, TRANSFERASE
422	1c0n	A	390	581	1.8e-16	0.13	0.19		CSDB PROTEIN; CHAIN: A;	LYASE ALPHA/BETA FOLD
422	1c7n	A	409	576	4e-09	0.05	-0.05		CRYSTALYSIN; CHAIN: A, B, C, D, E, F, G, H;	TRANSFERASE TRANSFERASE, AMINOTRANSFERASE, PYRIDOXAL PHOSPHATE
422	1c11	A	386	581	2e-25	0.30	0.54		CYSTATHIONINE BETA-LYASE; CHAIN: A, B;	METHIONINE BIOSYNTHESIS BETA CYSTATHIONASE; PLP-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
422	1cs1	A	406	564	6e-11	-0.15	0.59		CYSTATHIONINE GAMMA-SYNTHASE; CHAIN: A, B, C, D;	DEPENDENT ENZYMES, METHIONINE BIOSYNTHESIS, C-S BETA 2 LYASE
422	1d2f	A	409	588	1.4e-07	0.20	0.48		MALY PROTEIN; CHAIN: A, B;	LYASE CGS; LYASE, LLP-DEPENDENT ENZYMES, METHIONINE BIOSYNTHESIS
422	1dfo	A	389	532	1.2e-14	-0.45	0.40		SERINE HYDROXYMETHYLTRANSF ERASE; CHAIN: A, B, C, D;	TRANSFERASE AMINOTRANSFERASE FOLD, LARGE PLP-BINDING DOMAIN, SMALL C-2 TERMINAL DOMAIN, OPEN ALPHA-BETA STRUCTURE.
422	1eji	A	389	531	1.8e-06	-0.24	0.05		SERINE HYDROXYMETHYLTRANSF ERASE; CHAIN: A, B, C, D;	TRANSFERASE SHMT, SERINE METHYLASE; ALPHA PLP ASPARTATE, AMINO
422	1eju	A	416	566	1.4e-10	0.01	0.29		L-CYSTEINE/L-CYSTINE C-S LYASE; CHAIN: A, B;	TRANSFERASE (AAT)-LIKE FOLD
422	1etj	A	24	101	3.4e-11	0.31	0.01		TRANSCRIPTIONAL REPRESSOR TUP1; CHAIN: A, B, C;	TRANSFERASE SHMT, SERINE-GLYCINE CONVERSION, PYRIDOXAL 5'-PHOSPHATE, 2 TETRAHYDROFOLATE, ASYMMETRIC DIMER
422	1etj	A	28	306	1.7e-45	0.52	0.90		TRANSCRIPTIONAL REPRESSOR TUP1; CHAIN: A, B, C;	LYASE FES CLUSTER BIOSYNTHESIS, PYRIDOXAL 5'-PHOSPHATE, 2 THIOCYSTEINE, AMINOACRYLATE, ENZYME-PRODUCT COMPLEX
										TRANSCRIPTION INHIBITOR BETA-PROPELLER
										TRANSCRIPTION INHIBITOR BETA-PROPELLER

250

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
422	1erj	A	69	351	6.8e-49	0.48	0.99		TRANSCRIPTIONAL REPRESSOR TUP1; CHAIN: A, B, C;	TRANSCRIPTION INHIBITOR BETA-PROPELLER
422	1got	B	108	402	1e-29	0.20	0.16		GT-ALPHA/GI-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT; GAMMA1, TRANSDUCIN GAMMA SUBUNIT; COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
422	1got	B	19	245	5.1e-42	0.34	1.00		GT-ALPHA/GI-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT; GAMMA1, TRANSDUCIN GAMMA SUBUNIT; COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
422	1got	B	19	377	1e-56			75.34	GT-ALPHA/GI-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT; GAMMA1, TRANSDUCIN GAMMA SUBUNIT; COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
422	1got	B	39	349	1e-56	0.39	0.64		GT-ALPHA/GI-ALPHA CHIMERA; CHAIN: A; GT-BETA; CHAIN: B; GT-GAMMA; CHAIN: G;	COMPLEX (GTP-BINDING/TRANSDUCER) BETA1, TRANSDUCIN BETA SUBUNIT; GAMMA1, TRANSDUCIN GAMMA

251

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
422	1gkx	A	474	563	8e-08	-0.19	0.16		4-AMINO BUTYRATE AMINOTRANSFERASE; CHAIN: A, B, C, D;	SUBUNIT, COMPLEX (GTP-BINDING/TRANSDUCER), G PROTEIN, HETEROTRIMER 2 SIGNAL TRANSDUCTION
422	1qgn	A	386	579	4e-29	0.16	1.00		CYSTATHIONINE GAMMA-SYNTHASE; CHAIN: A, B, C, D, E, F, G, H;	TRANSFERASE GABA-AT; PLP-DEPENDENT ENZYME, AMINOTRANSFERASE, 4-AMINO BUTYRIC ACID, 2 ANTEPILEPTIC DRUG TARGET
422	1qj5	A	417	581	8e-19	0.13	0.24		7,8-DIAMINOPELARGONIC ACID SYNTHASE; CHAIN: A, B;	LYASE METHIONINE BIOSYNTHESIS, PYRIDOXAL 5'-PHOSPHATE, GAMMA-2 FAMILY, LYASE
422	1pjl	A	409	528	2e-06	-0.06	0.48		LYASE(CARBON-CARBON) TYROSINE PHENOL-LYASE (E.C.4.1.99.2) 1TPL 3	AMINOTRANSFERASE AMINOTRANSFERASE, PYRIDOXAL-5'-PHOSPHATE, BIOTIN 2 BIOSYNTHESIS
422	2oat	A	406	596	4e-19	0.14	0.04		ORNITHINE AMINOTRANSFERASE; CHAIN: A, B, C;	AMINOTRANSFERASE AMINOTRANSFERASE, 5-FLUOROMETHYLORNITHINE, PLP-DEPENDENT 2 ENZYME, PYRIDOXAL PHOSPHATE
422	2pjl	A	388	576	4e-28	0.07	0.00		TYROSINE PHENOL-LYASE; CHAIN: A, B;	LYASE LYASE, PLP-DEPENDENT ENZYME, PYRIDOXAL PHOSPHATE
424	1av1	A	172	363	3.4e-06			61.31	APOLIPOPROTEIN A-I; CHAIN: A, B, C, D;	LIPID TRANSPORT APO A-I; LIPOPROTEIN, LIPID TRANSPORT,



252

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
.										CHOLESTEROL METABOLISM, 2 ATHEROSCLEROSIS, HDL, LCAT- ACTIVATION
424	1cun	A	94	301	1e-05			67.32	ALPHA SPECTRIN; CHAIN: A, B, C;	STRUCTURAL PROTEIN TWO REPEATS OF SPECTRIN, ALPHA HELICAL LINKER REGION, 2 2 TANDEM 3-HELIX COILED-COILS, STRUCTURAL PROTEIN
424	1fk	A	231	319	8.5e-05	0.11	0.21		PREFOLDIN; CHAIN: A; PREFOLDIN; CHAIN: B; PREFOLDIN; CHAIN: C;	CHAPERONE ARCHAEAL PROTEIN
424	1quu	A	95	347	2e-10			61.81	HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2; CHAIN: A;	CONTRACTILE PROTEIN TRIPLE- HELIX COILED COIL, CONTRACTILE PROTEIN
430	1apm	E	1	337	0			71.47	TRANSFERASE(PHOSPHOTR ANSFERASE) SC-/AMPS- DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (SC/APKS) IAPM 3 (CATALYTIC SUBUNIT) "ALPHA" ISOENZYME MUTANT WITH SER 139 IAPM 4 REPLACED BY ALA (S139A) COMPLEX WITH THE PEPTIDE IAPM 5 INHIBITOR PKI(S-24) AND THE DETERGENT MEGA-8 IAPM 6	
430	1apm	E	6	342	0	0.37	0.80		TRANSFERASE(PHOSPHOTR ANSFERASE) SC-/AMPS-	

253

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (SC/APK\$) IAPM 3 (CATALYTIC SUBUNIT) "ALPHA" ISOENZYME MUTANT WITH SER 139 IAPM 4 REPLACED BY ALA (S139A\$) COMPLEX WITH THE PEPTIDE IAPM 5 INHIBITOR PKI(5-24) AND THE DETERGENT MEGA-8 IAPM 6	
430	1cki	A	1	294	1e-94	0.77	1.00		CASEIN KINASE I DELTA; ICKI 6 CHAIN: A, B; ICKI 7	PHOSPHOTRANSFERASE PROTEIN KINASE ICKI 18
430	1cki	A	1	298	1e-94			474.37	CASEIN KINASE I DELTA; ICKI 6 CHAIN: A, B; ICKI 7	PHOSPHOTRANSFERASE PROTEIN KINASE ICKI 18
430	1cmk	E	1	337	0			66.84	PHOSPHOTRANSFERASE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT 1CMK 3 (E.C.2.7.1.37) 1CMK 4	
430	1cmk	E	6	342	0	0.12	0.83		PHOSPHOTRANSFERASE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT 1CMK 3 (E.C.2.7.1.37) 1CMK 4	
430	1cp	E	1	315	0			73.32	TRANSFERASE (PHOSPHOTRANSFERASE) CAMP-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (CAPK) ICTP 3 (CATALYTIC	

254

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
430	1cnp	E	6	339	0	0.30	0.71		SUBUNIT) ICTP 4 TRANSFERASE(PHOSPHOTRANSFERASE) CAMP-DEPENDENT PROTEIN KINASE (E.C.2.7.1.37) (CAPK) ICTP 3 (CATALYTIC SUBUNIT) ICTP 4	
432	1chc		10	73	4e-09	-0.40	0.90		VIRUS EQUINE HERPES VIRUS-1 (C3HC4, OR RING DOMAIN) 1CHC 3 (NMR, 1 STRUCTURE) 1CHC 4	
432	1bv	A	13	49	6.8e-05	-0.51	0.04		SIGNAL TRANSDUCTION PROTEIN CBL; CHAIN: A; ZAP-70 PEPTIDE; CHAIN: B; UBIQUITIN-CONJUGATING ENZYME E12-18 KDA UBCH7; CHAIN: C;	LIGASE CBL, UBCH7, ZAP-70, E2, UBIQUITIN, E3, PHOSPHORYLATION, 2 TYROSINE KINASE, UBIQUITINATION, PROTEIN DEGRADATION,
432	1g25	A	10	62	2e-09	-0.38	0.41		CDK-ACTIVATING KINASE ASSEMBLY FACTOR MAT1; CHAIN: A;	METAL BINDING PROTEIN RING FINGER PROTEIN MAT1; RING FINGER (C3HC4)
432	1md		10	62	2e-09	-0.07	0.37		RAG1; CHAIN: NULL;	DNA-BINDING PROTEIN V(D)J RECOMBINATION ACTIVATING PROTEIN 1; RAG1, V(D)J RECOMBINATION, ANTIBODY, MAD, RING FINGER, 2 ZINC BINUCLEAR CLUSTER, ZINC FINGER, DNA-BINDING PROTEIN
432	1md		13	49	3.4e-05	-0.28	0.18		RAG1; CHAIN: NULL;	DNA-BINDING PROTEIN V(D)J RECOMBINATION ACTIVATING PROTEIN 1; RAG1, V(D)J

255

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
439	Imof		1	32	8.5e-13	-0.55	0.13		MOLONEY MURINE LEUKEMIA VIRUS P15; CHAIN: NULL;	RECOMBINATION, ANTIBODY, MAD, RING FINGER, 2 ZINC BINUCLEAR CLUSTER, ZINC FINGER, DNA-BINDING PROTEIN
449	ledh	A	129	332	6e-29	0.26	0.98		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	ledh	A	129	332	8.5e-29	0.19	0.94		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	ledh	A	17	228	1.7e-29	0.32	0.93		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	ledh	A	2	109	1.2e-27	0.36	0.77		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	ledh	A	242	442	4e-30			84.86	E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS

256

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
449	1edh	A	243	442	3.4e-27	0.20	0.58		E-CADHERIN; CHAIN: A, B;	1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	1edh	A	245	442	4e-30	0.29	0.99		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	1ncg		124	227	4e-05	-0.16	0.07		N-CADHERIN; INCG 3	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
449	1ncg		17	107	8.5e-07	0.06	0.09		N-CADHERIN; INCG 3	CELL ADHESION PROTEIN
449	1ncg		241	331	1.2e-14	0.46	0.21		N-CADHERIN; INCG 3	CELL ADHESION PROTEIN
449	1ncg		26	108	1.2e-09	0.66	0.42		N-CADHERIN; INCG 3	CELL ADHESION PROTEIN
449	1ncg		268	330	0.00017	-0.02	0.04		N-CADHERIN; INCG 3	CADHERIN INCG 13
449	1ncg		352	440	2e-08	0.23	0.10		N-CADHERIN; INCG 3	CELL ADHESION PROTEIN
449	1nci	B	230	332	6e-14	0.30	0.82		N-CADHERIN; INCI 3	CADHERIN INCG 13
449	1nci	B	276	332	0.00015	-0.59	0.66		N-CADHERIN; INCI 3	CELL ADHESION PROTEIN
449	1nci	B	368	442	1.8e-08	-0.06	0.13		N-CADHERIN; INCI 3	CADHERIN INCI 13

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
449	Inci	B	43	109	8.5e-07	0.24	0.09		N-CADHERIN; INCI 3	CELL ADHESION PROTEIN CADHERIN INCI 13
449	Inci	A	1	109	6.8e-33	-0.08	0.37		N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
449	Inci	A	127	334	8.5e-31			94.66	N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
449	Inci	A	129	332	8.5e-31	0.01	0.36		N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
449	Inci	A	17	228	1.4e-32	0.23	0.58		N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
449	Inci	A	243	442	3.4e-26	0.27	0.81		N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
449	Inci		125	232	2e-07	-0.22	0.07		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
449	Inci		129	232	1.7e-05	-0.60	0.01		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
449	Inci		17	113	1.7e-09	0.69	0.62		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
449	Inci		243	336	8.5e-07	0.23	0.33		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
449	Inci		245	336	1e-17	0.53	0.80		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
449	Inci		26	113	1e-10	0.52	0.03		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
449	Inci		354	442	4e-09	0.29	0.17		EPITHELIAL CADHERIN;	CELL ADHESION UVOMORULIN;

258

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CHAIN: NULL;	CADHERIN, CALCIUM BINDING, CELL ADHESION
453	1eut		537	744	2.4e-09	0.03	-0.20		SIALIDASE; CHAIN: NULL;	HYDROLASE NEURAMINIDASE; HYDROLASE, GLYCOSIDASE
453	2tbv	C	290	580	4.8e-14	0.02	-0.20		VIRUS TOMATO BUSHY STUNT VIRUS 2TBV 4	
462	1lal	A	13	282	0.00018	0.04	0.34		IMPORTIN ALPHA; CHAIN: A;	NUCLEAR IMPORT RECEPTOR KARYOPHERIN ALPHA; NUCLEAR IMPORT RECEPTOR, NUCLEAR LOCALIZATION SIGNAL, 2 ARMADILLO REPEATS, AUTOINHIBITION, INTRASTERIC REGULATION
462	1lal	A	2	439	0.002	0.02	0.06		IMPORTIN ALPHA; CHAIN: A;	NUCLEAR IMPORT RECEPTOR KARYOPHERIN ALPHA; NUCLEAR IMPORT RECEPTOR, NUCLEAR LOCALIZATION SIGNAL, 2 ARMADILLO REPEATS, AUTOINHIBITION, INTRASTERIC REGULATION
462	1lal	A	20	466	0.002			93.29	IMPORTIN ALPHA; CHAIN: A;	NUCLEAR IMPORT RECEPTOR KARYOPHERIN ALPHA; NUCLEAR IMPORT RECEPTOR, NUCLEAR LOCALIZATION SIGNAL, 2 ARMADILLO REPEATS, AUTOINHIBITION, INTRASTERIC REGULATION
462	1lbr	B	13	202	2e-05	-0.00	0.40		RAN; CHAIN: A, C; IMPORTIN BETA SUBUNIT;	SMALL GTPASE KARYOPHERIN BETA, P95 SMALL GTPASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
462	2bct		195	653	0.001	0.04	0.27		CHAIN: B, D; BETA-CATENIN; CHAIN: NULL;	NUCLEAR TRANSPORT RECEPTOR STRUCTURAL PROTEIN ARMADILLO REPEAT, BETA- CATENIN, STRUCTURAL PROTEIN
462	3bct		3	417	1e-09	-0.03	0.53		BETA-CATENIN; CHAIN: NULL;	ARMADILLO REPEAT ARMADILLO REPEAT, BETA- CATENIN, CYTOSKELETON
464	1atl	A	144	345	3.4e-39			149.86	ATROLYSIN C; 1ATL 4 CHAIN: A, B, C, D; 1ATL 5	METALLOENDOPEPTIDASE HEMORRHAGIC TOXIN C, FORM D; 1ATL 6
464	1atl	A	144	345	3.4e-39	0.97	1.00		ATROLYSIN C; 1ATL 4 CHAIN: A, B, C, D; 1ATL 5	METALLOENDOPEPTIDASE HEMORRHAGIC TOXIN C, FORM D; 1ATL 6
464	1bkc	A	144	343	2.4e-49	0.22	0.95		TUMOR NECROSIS FACTOR- ALPHA-CONVERTING ENZYME; CHAIN: A, C, E, I;	ZN-ENDOPEPTIDASE TACE; ZN- ENDOPEPTIDASE, HYDROLASE, TNF-ALPHA
464	1bud	A	145	343	1.2e-64			141.49	ACUTOLYSIN A; CHAIN: A;	TOXIN HEMORRHAGIN I, IAAH-I; METALLOPROTEINASE, SNAKE VENOM, MMP, TOXIN
464	1bud	A	145	343	8.5e-39	0.81	1.00		ACUTOLYSIN A; CHAIN: A;	TOXIN HEMORRHAGIN I, IAAH-I; METALLOPROTEINASE, SNAKE VENOM, MMP, TOXIN
464	1bud	A	146	343	1.2e-64	0.97	1.00		ACUTOLYSIN A; CHAIN: A;	TOXIN HEMORRHAGIN I, IAAH-I; METALLOPROTEINASE, SNAKE VENOM, MMP, TOXIN
464	1dx5	I	299	411	8.5e-09	0.14	-0.18		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
464	1enn		414	503	1.7e-10	0.32	0.00		FIBRILLIN; CHAIN: NULL;	MATRIX PROTEIN EXTRACELLULAR MATRIX, CALCIUM-BINDING, GLYCOPROTEIN, 2 REPEAT, SIGNAL, MULTIGENE FAMILY, DISEASE MUTATION, 3 EGF-LIKE DOMAIN, HUMAN FIBRILLIN-1 FRAGMENT, MATRIX PROTEIN
464	1enn		469	543	5.1e-12	0.22	-0.15		FIBRILLIN; CHAIN: NULL;	MATRIX PROTEIN EXTRACELLULAR MATRIX, CALCIUM-BINDING, GLYCOPROTEIN, 2 REPEAT, SIGNAL, MULTIGENE FAMILY, DISEASE MUTATION, 3 EGF-LIKE DOMAIN, HUMAN FIBRILLIN-1 FRAGMENT, MATRIX PROTEIN
464	1f5y	A	440	523	5.1e-08	0.18	-0.18		LOW-DENSITY LIPOPROTEIN RECEPTOR; CHAIN: A;	LIPID BINDING PROTEIN LDL RECEPTOR; BETA HAIRPIN, 3-10 HELIX, CALCIUM BINDING
464	1fv1		364	432	5.1e-14	0.25	1.00		FLAVORIDIN; IFVL 4 CHAIN: NULL, IFVL 5	BLOOD COAGULATION INHIBITOR GP IIB/IIIA ANTAGONIST IFVL 9
464	1iag		142	345	2.4e-65			140.05	METALLOPROTEASE ADAMALYSIN II (PROTEINASE II) (E.C.3.4.24.46) I1AG 3	
464	1iag		144	345	8.5e-39	0.90	1.00		METALLOPROTEASE	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									ADAMALYSIN II (PROTEINASE II) (E.C.3.4.24.46) IIAG 3	
464	1iag		146	345	2.4e-65	0.70	1.00		METALLOPROTEASE ADAMALYSIN II (PROTEINASE II) (E.C.3.4.24.46) IIAG 3	
464	1klo		343	496	6.8e-11	0.18	-0.19		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
464	1klo		460	603	5.1e-12	0.06	-0.19		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
464	1klo		591	744	3.4e-12	0.02	-0.19		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
464	1kst		364	432	1.5e-14	0.53	0.89		AGGREGATION INHIBITOR, GP ANTAGONIST KISTRIN (NMR, 8 STRUCTURES) 1KST 3	
464	1pfx	L	414	482	5.1e-08	-0.00	-0.19		FACTOR IXA; CHAIN: C, L; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR: COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
464	1pfx	L	681	739	5.1e-08	0.12	-0.20		FACTOR IXA; CHAIN: C, L; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR: COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
464	1qua	A	144	343	6.8e-38	0.92	1.00		ACUTOLYSIN-C; CHAIN: A;	TOXIN HEMORRHAGIN III

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
										METALLOPROTEASE, HEMORRHAGIC TOXIN, SNAKE VENOM PROTEINASE, 2 CRYSTAL STRUCTURE, AGKISTRODON ACUTUS
464	1qub	A	337	619	1.2e-09	0.08	-0.17		HUMAN BETA2-GLYCOPROTEIN I; CHAIN: A;	MEMBRANE ADHESION SHORT CONSENSUS REPEAT, SUSHI, COMPLEMENT CONTROL PROTEIN, 2 N-GLYCOSYLATION, MULTI-DOMAIN, MEMBRANE ADHESION
464	1tpg		571	617	8.5e-07	-0.63	0.10		T-PLASMINOGEN ACTIVATOR FI-G; 1TPG 7 CHAIN: NULL; 1TPG 8	PLASMINOGEN ACTIVATION
464	9wga	A	158	356	3.4e-11	0.06	-0.19		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
464	9wga	A	239	419	1.7e-12	0.10	-0.18		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
464	9wga	A	289	473	6.8e-14	0.38	-0.17		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
464	9wga	A	345	506	3.4e-14	0.16	-0.18		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
464	9wga	A	393	578	5.1e-12	0.23	-0.09		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
464	9wga	A	443	617	6.8e-13	0.11	-0.19		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									(ISOLECTIN 2) 9WGA 3	
468	1b3u	A	279	582	3.6e-06	0.14	0.96		PROTEIN PHOSPHATASE PP2A; CHAIN: A, B;	SCAFOLD PROTEIN SCAFFOLD PROTEIN, PP2A, PHOSPHORYLATION, HEAT REPEAT
468	1ec4	A	183	462	4.8e-05	0.40	0.24		KARYOPHERIN ALPHA; CHAIN: A, B; MYC PROTO-ONCOGENE PROTEIN; CHAIN: C, D, E, F;	TRANSPORT PROTEIN SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN; ARM REPEAT
468	1ec4	A	246	580	1.2e-09	0.35	0.51		KARYOPHERIN ALPHA; CHAIN: A, B; MYC PROTO-ONCOGENE PROTEIN; CHAIN: C, D, E, F;	TRANSPORT PROTEIN SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN; ARM REPEAT
468	1qgr	A	102	580	3.6e-11	0.02	0.12		IMPORTIN BETA SUBUNIT; CHAIN: A; IMPORTIN ALPHA-2 SUBUNIT; CHAIN: B;	TRANSPORT RECEPTOR KARYOPHERIN BETA-1, NUCLEAR FACTOR P97, IMPORTIN IMPORTIN ALPHA-2 SUBUNIT, KARYOPHERIN ALPHA-2 TRANSPORT RECEPTOR, NUCLEAR IMPORT, HEAT MOTIF, NLS-BINDING
468	3bct		158	579	8.4e-16			93.61	BETA-CATENIN; CHAIN: NULL;	ARMADILLO REPEAT ARMADILLO REPEAT, BETA-CATENIN, CYTOSKELETON
468	3bct		198	580	8.4e-16	0.41	0.99		BETA-CATENIN; CHAIN: NULL;	ARMADILLO REPEAT ARMADILLO REPEAT, BETA-CATENIN, CYTOSKELETON
476	1a17		653	765	1.2e-07	-0.02	0.13		SERINE/THREONINE PROTEIN PHOSPHATASE 5;	HYDROLASE TETRATRICOPEPTIDE, TRP;

264

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									CHAIN: NULL;	HYDROLASE, PHOSPHATASE, PROTEIN-PROTEIN INTERACTIONS, TPR, 2 SUPER-HELIX, X-RAY STRUCTURE
476	1elr	A	623	726	3.4e-08	0.07	-0.20		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE MEEVD; CHAIN: B;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSP90, 2 PROTEIN BINDING
476	1elr	A	66	193	1.7e-22	0.08	0.30		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE MEEVD; CHAIN: B;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSP90, 2 PROTEIN BINDING
476	1elr	A	662	772	3.4e-11	0.27	0.23		TPR2A-DOMAIN OF HOP; CHAIN: A; HSP90-PEPTIDE MEEVD; CHAIN: B;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSP90, 2 PROTEIN BINDING
476	1elw	A	657	784	1.7e-09	-0.08	0.70		TPR1-DOMAIN OF HOP; CHAIN: A, B; HSC70-PEPTIDE; CHAIN: C, D;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSC70, 2 HSP70, PROTEIN BINDING
476	1elw	A	66	183	6.8e-23	-0.10	0.62		TPR1-DOMAIN OF HOP; CHAIN: A, B; HSC70-PEPTIDE; CHAIN: C, D;	CHAPERONE HOP, TPR-DOMAIN, PEPTIDE-COMPLEX, HELICAL REPEAT, HSC70, 2 HSP70, PROTEIN BINDING
476	1fch	A	21	207	3.4e-19	-0.12	0.03		PEROXISOMAL TARGETING SIGNAL 1 RECEPTOR; CHAIN: A, B; PTS1-CONTAINING PEPTIDE; CHAIN: C, D;	SIGNALING PROTEIN PEROXISOMAL RECEPTOR 1, PTS1-BP, PEROXIN-5, PTS1 PROTEIN-PEPTIDE COMPLEX, TETRATRICOPEPTIDE REPEAT, TPR, 2 HELICAL REPEAT
476	1fch	A	639	801	1.5e-14	0.22	0.57		PEROXISOMAL TARGETING	SIGNALING PROTEIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									SIGNAL 1 RECEPTOR; CHAIN: A, B; PTS1-CONTAINING PEPTIDE; CHAIN: C, D;	PEROXISOME RECEPTOR 1, PTS1-BP, PEROXIN-5, PTS1 PROTEIN-PEPTIDE COMPLEX, TETRAPEPTIDE REPEAT, TPR, 2 HELICAL REPEAT
481	1bzk	A	396	436	6e-16	-0.88	0.05		BAND 3 ANION TRANSPORT PROTEIN; CHAIN: A;	TRANSPORT PROTEIN HUMAN ERYTHROCYTE ANION TRANSPORTER, TRANSMEMBRANE, 2 SYNTHETIC PEPTIDE, NMR
482	1awc	B	173	364	2.2e-17	0.32	0.99		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
482	1awc	B	271	445	3.4e-34	-0.04	0.51		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
482	1bd8		151	320	3.4e-25	-0.42	0.06		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF

266

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
482	1bd8		173	367	1.1e-17	0.08	0.94		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
482	1bd8		271	445	1.5e-28	0.06	0.88		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
482	1bi7	B	488	604	1e-21	0.10	-0.20		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A, MTS1; CYCLIN DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
482	1bx	B	219	368	2.2e-16	0.25	0.87		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
482	1bx	B	271	445	3.4e-27	-0.22	0.55		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
482	1bu9	A	219	368	2e-17	0.05	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
482	1bu9	A	271	450	3.4e-28	-0.23	0.25		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR HORMONE/GROWTH FACTOR P18- INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
482	1bu9	A	433	609	1.7e-31	0.12	-0.19		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18- INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR SUPPRESSOR, CYCLIN-2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
482	1lhb	A	219	368	6.6e-17	0.03	0.95		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18- INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANK YRIN REPEAT, 2 CDK 4/6 INHIBITOR
482	1lhb	A	271	449	1.4e-27	-0.18	0.58		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18- INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANK YRIN REPEAT, 2 CDK 4/6 INHIBITOR
482	1lkn	D	131	368	2.2e-19	-0.16	0.84		NF-KAPPA-B P65 SUBUNIT; CHAIN: A; NF-KAPPA-B P50D SUBUNIT; CHAIN: C; I- KAPPA-B-ALPHA; CHAIN: D;	TRANSCRIPTION FACTOR P65; P50D; TRANSCRIPTION FACTOR, IKB/NFKB COMPLEX
482	1lkn	D	172	409	3.4e-36	-0.16	0.01		NF-KAPPA-B P65 SUBUNIT; CHAIN: A; NF-KAPPA-B P50D SUBUNIT; CHAIN: C; I- KAPPA-B-ALPHA; CHAIN: D;	TRANSCRIPTION FACTOR P65; P50D; TRANSCRIPTION FACTOR, IKB/NFKB COMPLEX
482	1myo		152	288	1.7e-14	-0.38	0.01		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-



Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
482	1nf	E	171	409	1.7e-36	-0.09	0.18		NF-KAPPA-B P65; CHAIN: A, C; NF-KAPPA-B P50; CHAIN: B, D; I-KAPPA-B-ALPHA; CHAIN: E, F;	REPEAT COMPLEX (TRANSCRIPTION REG/ANK REPEAT) COMPLEX (TRANSCRIPTION REGULATION/ANK REPEAT), ANKYRIN 2 REPEAT HELIX
482	1sw6	A	217	387	6.8e-17	-0.52	0.00		REGULATORY PROTEIN SWI6; CHAIN: A, B;	TRANSCRIPTION REGULATION, ANKYRIN REPEATS, CELL-CYCLE
482	1sw6	A	96	306	4.4e-09	-0.01	0.89		REGULATORY PROTEIN SWI6; CHAIN: A, B;	TRANSCRIPTION REGULATION, ANKYRIN REPEATS, CELL-CYCLE
485	1cvs	D	76	272	8.5e-29	0.05	-0.07		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B; FIBROBLAST GROWTH FACTOR RECEPTOR 1; CHAIN: C, D;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF, FGFR, IMMUNOGLOBULIN-LIKE, SIGNAL TRANSDUCTION, 2 DIMERIZATION, GROWTH FACTOR/GROWTH FACTOR RECEPTOR
485	1epf	A	66	257	3.4e-19	-0.01	0.10		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM, NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
485	1ev2	E	76	272	5.1e-24	0.08	-0.05		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C, D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2, FGFR2, IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
485	1ev2	G	76	275	1.4e-27	-0.27	0.23		FIBROBLAST GROWTH FACTOR 2; CHAIN: A, B, C;	GROWTH FACTOR/GROWTH FACTOR RECEPTOR FGF2, FGFR2;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									D; FIBROBLAST GROWTH FACTOR RECEPTOR 2; CHAIN: E, F, G, H;	IMMUNOGLOBULIN (IG) LIKE DOMAINS BELONGING TO THE I-SET 2 SUBGROUP WITHIN IG-LIKE DOMAINS, B-TREFOIL FOLD
485	1hg	A	60	163	1.7e-17	0.44	0.58		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
485	1ncl		64	164	6.8e-16	0.09	0.41		TTTN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN, NEXTM5; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN FOLD, ALTERNATIVE SPLICING, SIGNAL, 3 MUSCLE PROTEIN
485	1nm		66	164	1e-15	0.34	0.33		MUSCLE PROTEIN TITIN MODULE M5 (CONNECTIN) 1TNM 3 (NMR, MINIMIZED AVERAGE STRUCTURE) 1TNM 4 1TNM 58	
493	1a4y	A	74	224	1.1e-15	0.16	0.15		RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE), COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPTOPE MAPPING, LEUCINE-RICH 3 REPEATS
493	1a9n	A	101	235	4.4e-24	0.30	0.48		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
493	1a9n	A	74	211	4.4e-18	0.22	0.69		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	SNRNP, RIBONUCLEOPROTEIN COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
493	1a9n	C	101	235	1.5e-23	0.46	0.53		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
493	1a9n	C	74	211	6.6e-18	0.38	0.96		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A; CHAIN: A, C; U2 B; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP, RIBONUCLEOPROTEIN
493	1d0b	A	6	181	3.4e-20	0.34	0.10		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
493	1d0b	A	69	268	1.2e-19	-0.00	0.53		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
493	2bnh		81	230	8.8e-22	0.28	-0.03		RIBONUCLEASE INHIBITOR; CHAIN: NULL;	ACETYLATION RNASE INHIBITOR, RIBONUCLEASE/ANGIOGENIN INHIBITOR ACETYLATION, LEUCINE-RICH REPEATS
503	1awc		513	600	3.4e-09	0.06	-0.15		SOS1; CHAIN: NULL;	SIGNAL TRANSDUCTION SIGNAL TRANSDUCTION, SOS, PLECKSTRIN HOMOLOG (PH) DOMAIN
503	1by1	A	308	520	4e-44	0.02	1.00		PIX; CHAIN: A;	TRANSPORT PROTEIN RHO-GTPASE EXCHANGE FACTOR, TRANSPORT PROTEIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
503	1by1	A	320	483	1.2e-31	-0.20	0.99		PIX; CHAIN: A;	TRANSPORT PROTEIN RHO-GTPASE EXCHANGE FACTOR, TRANSPORT PROTEIN
503	1dbh	A	314	600	8.5e-31	0.17	0.92		HUMAN SOS 1; CHAIN: A;	GENE REGULATION SON OF SEVENTEENLESS PROTEIN; GUANINE NUCLEOTIDE EXCHANGE FACTOR, GENE REGULATION
503	1dbh	A	317	613	2e-43	0.00	0.86		HUMAN SOS 1; CHAIN: A;	GENE REGULATION SON OF SEVENTEENLESS PROTEIN; GUANINE NUCLEOTIDE EXCHANGE FACTOR, GENE REGULATION
503	1etj	A	900	1125	0.006	-0.05	0.51		TRANSCRIPTIONAL REPRESSOR TUP1; CHAIN: A, B, C;	TRANSCRIPTION INHIBITOR BETA-PROPELLER
503	1f5x	A	299	505	6e-36	0.39	1.00		RHO-GEF VAV; CHAIN: A;	SIGNALING PROTEIN 11 ALPHA-HELICES
503	1f5x	A	312	483	1.5e-33	0.25	1.00		RHO-GEF VAV; CHAIN: A;	SIGNALING PROTEIN 11 ALPHA-HELICES
507	1a0j	A	218	334	3.4e-42	0.34	0.29		TRYPSIN; CHAIN: A, B, C, D;	SERINE PROTEASE SERINE PROTEINASE, TRYPSIN, HYDROLASE
507	1a0j	A	351	566	1.7e-79			146.48	TRYPSIN; CHAIN: A, B, C, D;	SERINE PROTEASE SERINE PROTEINASE, TRYPSIN, HYDROLASE
507	1a0j	A	368	566	1.7e-79	0.40	1.00		TRYPSIN; CHAIN: A, B, C, D;	SERINE PROTEASE SERINE PROTEINASE, TRYPSIN, HYDROLASE
507	1a0l	A	315	566	5.1e-69			159.04	BETA-TRYPTASE; CHAIN: A, B, C, D;	SERINE PROTEINASE TRYPSIN-LIKE SERINE PROTEINASE, TETRAMER, HEPARIN, ALLERGY,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	1a5i	A	313	564	3.4e-70			147.63	PLASMINOGEN ACTIVATOR; CHAIN: A; GLU-GLY-ARG CHLOROMETHYL KETONE; CHAIN: I;	2 ASTHMA COMPLEX (SERINE PROTEASE/INHIBITOR) (DELTA FEK) DSPALPHA1; EGRCMK; SERINE PROTEASE, FIBRINOLYTIC ENZYMES, PLASMINOGEN 2 ACTIVATORS
507	1aks	A	218	323	3.4e-41	0.23	0.04		ALPHA TRYPSIN; CHAIN: A, B;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE
507	1aks	B	469	562	6.6e-36	0.19	1.00		ALPHA TRYPSIN; CHAIN: A, B;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE
507	1aut	C	330	564	3.4e-67			132.80	ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE, PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
507	1bio		380	562	6.6e-61	0.52	1.00		COMPLEMENT FACTOR D; CHAIN: NULL;	SERINE PROTEASE SERINE PROTEASE, HYDROLASE, COMPLEMENT, FACTOR D, CATALYTIC 2 TRIAD, SELF-REGULATION
507	1bru	P	218	333	1.7e-39	0.04	0.17		ELASTASE; CHAIN: P;	SERINE PROTEASE PPE; SERINE PROTEASE, HYDROLASE
507	1bru	P	318	566	3.4e-70			127.57	ELASTASE; CHAIN: P;	SERINE PROTEASE PPE; SERINE PROTEASE, HYDROLASE
507	1chg		336	565	6.8e-59			139.58	HYDROLASE ZYMOGEN (SERINE PROTEINASE) CHYMOTRYPSINOGEN A	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	1dan	H	335	568	6.8e-64			127.53	1CHG4 BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE- ARG- CHLOROMETHYLKETONE (DEFRCK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO- FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
507	1dpo		218	333	1.5e-42	0.11	0.23		TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, ZYMOGEN, 2 SIGNAL, MULTIGENE FAMILY
507	1dpo		330	566	6.8e-73			137.74	TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, ZYMOGEN, 2 SIGNAL, MULTIGENE FAMILY
507	1ekb	B	330	564	1.7e-72			170.52	ENTEROPEPTIDASE; CHAIN: A; ENTEROPEPTIDASE; CHAIN: B; VAL-ASP-ASP- ASP-ASP-LYS PEPTIDE; CHAIN: C;	HYDROLASE/HYDROLASE INHIBITOR ENTEROKINASE, HEAVY CHAIN; ENTEROKINASE, LIGHT CHAIN; ENTEROPEPTIDASE, TRYPSINOGEN ACTIVATION, 2 HYDROLASE/HYDROLASE INHIBITOR
507	1ekb	B	368	561	1.7e-72	0.44	1.00		ENTEROPEPTIDASE; CHAIN: A; ENTEROPEPTIDASE; CHAIN: B; VAL-ASP-ASP- ASP-ASP-LYS PEPTIDE; CHAIN: C;	HYDROLASE/HYDROLASE INHIBITOR ENTEROKINASE, HEAVY CHAIN; ENTEROKINASE, LIGHT CHAIN; ENTEROPEPTIDASE, TRYPSINOGEN ACTIVATION, 2 HYDROLASE/HYDROLASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	1fxy	A	330	567	6.8e-76			151.71	COAGULATION FACTOR XA-TRYPSIN CHIMERA; CHAIN: A; D-PHE-PRO-ARG-CHLOROMETHYLKETONE (PPACK) WITH CHAIN: I;	INHIBITOR
507	1fxy	A	367	566	6.8e-76	0.38	1.00		COAGULATION FACTOR XA-TRYPSIN CHIMERA; CHAIN: A; D-PHE-PRO-ARG-CHLOROMETHYLKETONE (PPACK) WITH CHAIN: I;	COMPLEX (PROTEASE/INHIBITOR) TRYPSIN, COAGULATION FACTOR XA, CHIMERA, PROTEASE, PPACK, 2 CHLOROMETHYLKETONE, COMPLEX (PROTEASE/INHIBITOR)
507	1gct	A	318	566	1.7e-62			143.21	HYDROLASE (SERINE PROTEINASE) GAMMA-CHYMOTRYPSIN *A (E.C.3.4.21.1) (SP*H 7.0) 1GCT 3	COMPLEX (PROTEASE/INHIBITOR) RTAP; GLYCOPROTEIN, SERINE PROTEASE, PLASMA, BLOOD COAGULATION, 2 COMPLEX (PROTEASE/INHIBITOR)
507	1kgg	H	330	567	6.6e-65			126.65	FACTOR XA; CHAIN: H, L; ANTICOAGULANT PEPTIDE; CHAIN: I;	COMPLEX (PROTEASE/INHIBITOR) RTAP; GLYCOPROTEIN, SERINE PROTEASE, PLASMA, BLOOD COAGULATION, 2 COMPLEX (PROTEASE/INHIBITOR)
507	1kgg	H	380	564	6.6e-65	0.43	1.00		FACTOR XA; CHAIN: H, L; ANTICOAGULANT PEPTIDE; CHAIN: I;	COMPLEX (PROTEASE/INHIBITOR) RTAP; GLYCOPROTEIN, SERINE PROTEASE, PLASMA, BLOOD COAGULATION, 2 COMPLEX (PROTEASE/INHIBITOR)
507	1mct	A	218	333	5.1e-44	0.17	0.15		COMPLEX (PROTEINASE/INHIBITOR) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH INHIBITOR FROM BITTER 1MCT 3 GOURD	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	1mct	A	337	566	3.4e-80			146.55	IMCT 4 COMPLEX(PROTEINASE/INHIBITOR) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH INHIBITOR FROM BITTER IMCT 3 GOURD IMCT 4	
507	1mct	A	368	566	3.4e-80	0.25	1.00		COMPLEX(PROTEINASE/INHIBITOR) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH INHIBITOR FROM BITTER IMCT 3 GOURD IMCT 4	
507	1mkx	K	290	564	1e-62			132.00	ALPHA-THROMBIN; CHAIN: L, H; PRETHROMBIN-2; CHAIN: K;	COMPLEX (BLOOD COAGULATION/PROENZYME) COMPLEX (BLOOD COAGULATION/PROENZYME), THROMBIN, 2 PRETHROMBIN-2, PLASMA, SERINE PROTEASE
507	1nrm	A	335	564	5.1e-65			131.31	NEUROPSIN; CHAIN: A, B;	SERINE PROTEINASE SERINE PROTEINASE, GLYCOPROTEIN
507	1pyt	C	380	562	2.2e-62	0.16	1.00		PROCARBOXYPEPTIDASE A; CHAIN: A, B; PROPROTEINASE E; CHAIN: C; CHYMOTRYPSINOGEN C; CHAIN: D;	TERNARY COMPLEX (ZYMOGEN) TC, PCPA-TG; TERNARY COMPLEX (ZYMOGEN), SERINE PROTEINASE, C-TERMINAL 2 PEPTIDASE
507	1pyt	D	314	564	2.2e-64			140.04	PROCARBOXYPEPTIDASE A; CHAIN: A, B; PROPROTEINASE E; CHAIN: C; CHYMOTRYPSINOGEN C; CHAIN: D;	TERNARY COMPLEX (ZYMOGEN) TC, PCPA-TG; TERNARY COMPLEX (ZYMOGEN), SERINE PROTEINASE, C-TERMINAL 2 PEPTIDASE



276

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	1pyt	D	380	562	2.2e-64	0.33	1.00		PROCARBOXYPEPTIDASE A; CHAIN: A, B; PROTEINASE E; CHAIN: C; CHYMOTRYPSINOGEN C; CHAIN: D;	TERNARY COMPLEX (ZYMOGEN) TC, PCPA-TC; TERNARY COMPLEX (ZYMOGEN), SERINE PROTEINASE, C-TERMINAL 2 PEPTIDASE
507	1qrz	A	312	566	8.5e-73			157.79	PLASMINOGEN; CHAIN: A, B, C, D;	HYDROLASE MICROPLASMINOGEN, SERINE PROTEASE, ZYMOGEN, CHYMOTRYPSIN 2 FAMILY, HYDROLASE
507	1qrz	A	367	566	8.5e-73	0.44	1.00		PLASMINOGEN; CHAIN: A, B, C, D;	HYDROLASE MICROPLASMINOGEN, SERINE PROTEASE, ZYMOGEN, CHYMOTRYPSIN 2 FAMILY, HYDROLASE
507	1rtf	B	330	565	3.4e-71			149.34	TWO CHAIN TISSUE PLASMINOGEN ACTIVATOR; CHAIN: A, B;	SERINE PROTEASE (TC)-T-PA; SERINE PROTEASE, FIBRINOLYTIC ENZYMES
507	1sgf	G	218	333	3.4e-37	-0.10	0.30		NERVE GROWTH FACTOR; CHAIN: A, B, G, X, Y, Z;	GROWTH FACTOR 7S NGF; GROWTH FACTOR (BETA-NGF), HYDROLASE - SERINE PROTEINASE 2 (GAMMA-NGF), INACTIVE SERINE PROTEINASE (ALPHA-NGF)
507	1sgf	G	330	567	5.1e-70			133.00	NERVE GROWTH FACTOR; CHAIN: A, B, G, X, Y, Z;	GROWTH FACTOR 7S NGF; GROWTH FACTOR (BETA-NGF), HYDROLASE - SERINE PROTEINASE 2 (GAMMA-NGF), INACTIVE SERINE PROTEINASE (ALPHA-NGF)
507	1slw	B	218	333	8.5e-40	-0.05	0.30		ECOTIN; CHAIN: A; ANIONIC	COMPLEX (SERINE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
									TRYPSIN; CHAIN: B;	PROTEASE/INHIBITOR) TRYPSIN INHIBITOR; SERINE PROTEASE, INHIBITOR, COMPLEX, METAL BINDING SITES, 2 PROTEIN ENGINEERING, PROTEASE-SUBSTRATE INTERACTIONS, 3 METALLOPROTEINS
507	1slw	B	351	566	3.4e-74			136.38	ECOTIN; CHAIN: A; ANIONIC TRYPSIN; CHAIN: B;	COMPLEX (SERINE PROTEASE/INHIBITOR) TRYPSIN INHIBITOR; SERINE PROTEASE, INHIBITOR, COMPLEX, METAL BINDING SITES, 2 PROTEIN ENGINEERING, PROTEASE-SUBSTRATE INTERACTIONS, 3 METALLOPROTEINS
507	1slw	B	368	566	3.4e-74	0.14	1.00		ECOTIN; CHAIN: A; ANIONIC TRYPSIN; CHAIN: B;	COMPLEX (SERINE PROTEASE/INHIBITOR) TRYPSIN INHIBITOR; SERINE PROTEASE, INHIBITOR, COMPLEX, METAL BINDING SITES, 2 PROTEIN ENGINEERING, PROTEASE-SUBSTRATE INTERACTIONS, 3 METALLOPROTEINS
507	1tm	A	218	333	8.5e-42	0.12	0.04		HYDROLASE (SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH THE INHIBITOR ITRN 3 DIISOPROPYL-FLUOROPHOSPHOFLUORIDATE (DFP) ITRN 4 HUMAN TRYPSIN, DFP INHIBITED	

278

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	1tm	A	351	567	5.1e-77			144.14	1TRN 6 HYDROLASE (SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH THE INHIBITOR 1TRN 3 DIISOPROPYL- FLUOROPHOSPHOFLUORID ATE (DFP) 1TRN 4 HUMAN TRYPSIN, DFP INHIBITED 1TRN 6	
507	1tm	A	368	566	5.1e-77	0.39	1.00		HYDROLASE (SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH THE INHIBITOR 1TRN 3 DIISOPROPYL- FLUOROPHOSPHOFLUORID ATE (DFP) 1TRN 4 HUMAN TRYPSIN, DFP INHIBITED 1TRN 6	
507	1try		380	563	2e-63	0.52	1.00		TRYPSIN; 1TRY 4 CHAIN: NULL; 1TRY 5	HYDROLASE (SERINE PROTEINASE)
507	2tbs		218	324	1.7e-39	-0.04	0.06		HYDROLASE(SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH BENZAMIDINE INHIBITOR 2TBS 3	
507	2tbs		351	566	3.4e-76			145.15	HYDROLASE(SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH BENZAMIDINE INHIBITOR 2TBS 3	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQ FOLD score	Compound	PDB annotation
507	2tbs		368	565	3.4e-76	0.37	1.00		HYDROLASE/SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH BENZAMIDINE INHIBITOR 2TBS 3	
507	5ppp		218	334	3.4e-42	0.28	0.09		BETA TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, 2 ZYMOGEN, SIGNAL
507	5ppp		351	566	3.4e-76			146.40	BETA TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, 2 ZYMOGEN, SIGNAL
507	5ppp		368	566	3.4e-76	0.36	1.00		BETA TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, 2 ZYMOGEN, SIGNAL
519	1aut	L	146	226	1e-08	0.05	-0.20		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
519	1emm		137	210	6.8e-15	0.28	-0.19		FIBRILLIN; CHAIN: NULL;	MATRIX PROTEIN EXTRACELLULAR MATRIX, CALCIUM-BINDING, GLYCOPROTEIN, 2 REPEAT, SIGNAL, MULTIGENE FAMILY, DISEASE MUTATION, 3 EGF-LIKE DOMAIN, HUMAN FIBRILLIN-1 FRAGMENT, MATRIX PROTEIN

280

Table 6

SEQ ID NO:	Position of Signal in Amino Acid Sequence	maxS (Maximum score)	means (Mean score)
277	34	0.972	0.868
278	34	0.972	0.868
279	34	0.972	0.868
280	17	0.994	0.966
281	28	0.983	0.868
282	37	0.997	0.957
283	16	0.917	0.844
284	31	0.931	0.621
285	22	0.972	0.883
286	40	0.972	0.632
287	34	0.964	0.760
288	49	0.936	0.594
289	19	0.952	0.897
290	26	0.914	0.727
291	27	0.911	0.682
292	22	0.996	0.941
293	24	0.986	0.955
294	25	0.938	0.818
295	32	0.969	0.872
296	32	0.986	0.926
297	16	0.971	0.564
298	23	0.982	0.801
299	28	0.995	0.945
300	27	0.908	0.613
301	22	0.981	0.771
302	19	0.958	0.722
304	32	0.983	0.825
305	21	0.991	0.897
306	20	0.990	0.957
307	24	0.948	0.690
308	36	0.959	0.788
309	41	0.979	0.594
310	34	0.943	0.677
311	24	0.974	0.934
312	24	0.974	0.882
313	31	0.952	0.767
314	18	0.956	0.868
315	18	0.956	0.868
316	24	0.910	0.559
317	30	0.992	0.941
318	25	0.989	0.809
319	40	0.971	0.570
321	32	0.967	0.612
322	21	0.913	0.732
323	40	0.945	0.778
324	28	0.949	0.828
325	49	0.987	0.628
326	19	0.990	0.910
327	39	0.996	0.766
328	39	0.996	0.766
329	39	0.996	0.766
330	42	0.988	0.594
331	49	0.976	0.581
332	28	0.959	0.747

281

Table 6

SEQ ID NO:	Position of Signal in Amino Acid Sequence	maxS (Maximum score)	means (Mean score)
333	26	0.934	0.688
334	36	0.959	0.880
335	24	0.961	0.843
336	34	0.929	0.666
337	32	0.984	0.903
338	42	0.970	0.642
339	42	0.970	0.642
340	37	0.969	0.747
341	25	0.983	0.861
342	43	0.979	0.635
343	20	0.990	0.944
344	49	0.981	0.658
345	24	0.984	0.915
346	24	0.984	0.878
347	26	0.982	0.899
348	41	0.959	0.578
349	21	0.947	0.760
350	23	0.908	0.781
351	39	0.997	0.792
352	32	0.971	0.794
353	36	0.978	0.716
354	16	0.992	0.973
355	16	0.990	0.967
356	35	0.988	0.849
357	25	0.936	0.710
358	49	0.993	0.675
359	44	0.993	0.648
360	44	0.994	0.700
361	36	0.966	0.818
362	19	0.981	0.958
363	42	0.991	0.608
364	25	0.958	0.613
365	30	0.883	0.630
366	49	0.971	0.749
367	29	0.977	0.879
368	48	0.995	0.760
369	22	0.972	0.883
370	17	0.983	0.915
443	21	0.899	0.686
489	39	0.925	0.610

282

Table 7

SEQ ID	Chromosomal location
1	3
2	3
3	3
4	17
5	1
8	4
9	22
10	1
11	1q32
12	15q21
13	10
14	4p15.1-p14
15	8
16	2q21-q22
18	6
19	X
22	12q24
23	17
24	4
26	8p22-q21.13
27	6q22.1-22.33
00001456Fb082	
28	1
29	5
30	5
31	6q22.2-22.33
33	11
35	11p15.5
36	19q13
37	19
38	17
39	17
40	2
41	4
42	20
44	7
46	12q
47	10
49	11
50	10
54	13
55	X
56	11q14
57	4
58	2
60	16p13.3
61	16q23.1
62	15q24
63	15q24
64	Xq13.1
66	4
67	16
68	11
69	19
70	19

283

Table 7

SEQ ID	Chromosomal location
71	10
72	9
73	6
74	4
75	9
79	11q13
80	5
82	1
83	1
84	11
85	17
90	1
91	19
92	19
93	22
94	6
96	18p11.2
97	3pter-3p25.1
98	1
99	18
100	18
101	15
102	15
103	17q21.2
106	22
108	15
109	10
110	10
112	10
113	11
114	2
116	5
117	4
118	5
119	10
120	22q13.1-13.33
121	13
122	20q13.11-13.2
123	6q13-14.3
124	3p21.2-p14.3
125	9q22.2-31.1
127	6
128	8
129	11
131	6
132	16
133	16
134	18
135	1
136	2
137	12
141	6
146	14
148	2
149	3q



284  
Table 7

SEQ ID	Chromosomal location
151	17
152	17
153	3p21
154	10
155	6
156	9q32-33.2
159	17
161	2
162	4
163	9
164	8
165	8
166	8
167	10
170	13
171	4
172	1
173	10
175	4q22-q24
178	20pter-q12
179	6
180	5q11
181	6p21.32-22.1
183	8q22
186	8
188	20p
189	19
190	19q13.4
192	8
194	20
195	1p12-13.2
196	6pter-p24.1
197	6pter-p24.1
199	8
200	17
201	19
202	19
203	19
204	1
205	5
207	9
208	21q11
209	4
210	12
211	14
212	19
213	9
215	1
216	15q14
218	Xq28
219	12
220	5q23
221	12q
222	16
223	20

285

Table 7

SEQ ID	Chromosomal location
225	2
226	3p
227	6
228	5
229	19
230	16
231	17
232	10
233	10
234	15
235	19
236	3p21.3
237	11
239	2
240	15
244	5
245	12q21.3-q21.4
246	17
247	3
248	20
249	15
250	7
251	6p12.3-21.1
252	8
253	4
254	3
255	10
256	19
257	19
258	19
259	16pter-p13
260	16pter-p13
262	9p13.1-13.3
263	19
265	16
266	7q22
269	15q14
270	11
271	11q23
272	X
273	11q12
274	3
275	2q23-q24
276	5

286

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score			
277	2	15-34; 1045	171-185; 1944		
278	2	15-34; 1045	147-161; 1944		
279	2	15-34; 1045	189-203; 1944		
280	6	42-58; 666	76-94; 864	119-136; 871	145-162; 929
		188-210; 1170	223-247; 1433		
281	2	43-65; 1330	104-119; 1947		
282	2	18-42; 2872	143-158; 1292		
283	8	21-48; 787	73-92; 1024	95-114; 1804	167-182; 1499
		210-225; 997	256-275; 1133	314-345; 939	389-405; 1337
284	9	16-32; 1965	40-59; 506	66-86; 2091	111-126; 1647
		155-172; 669	199-217; 1521	240-255; 1130	302-314; 951
			399-414; 2605		
285	5	576-592; 578	754-769; 2335	771-793; 1265	811-832; 1715
		863-878; 1373			
286	11	24-40; 2230	53-70; 1120	84-99; 2458	107-122; 1250
		144-160; 1641	221-237; 961	305-320; 1305	347-362; 1022
			380-398; 2785	400-415; 1417	466-487; 2904
287	2	16-31; 1313	314-336; 3340		
288	2	26-42; 1404	71-88; 2248		
289	1	36-54; 2289			
290	1	371-390; 2292			
291	4	14-33; 887	59-75; 2149	89-104; 1046	152-170; 547
292	2	70-87; 742	123-139; 630		
293	2	82-97; 1433	120-141; 1650		
294	1	200-221; 2645			
295	4	9-31; 1859	208-227; 607	394-414; 1433	469-491; 775
296	11	55-72; 1655	85-99; 938	123-138; 1548	242-254; 897
		284-303; 2550	347-363; 1621	381-401; 1905	430-445; 902470-484; 1799
			514-540; 888	559-574; 2224	
297	5	29-45; 1401	82-100; 1251	143-163; 2820	201-216; 1686
		228-251; 831			
298	8	40-62; 634	84-99; 2577	114-133; 1654	185-201; 2433
		228-245; 1509	328-346; 2079	414-432; 1097	434-451; 1182
299	4	68-84; 2529	77-112; 1338	98-120; 2138	147-182; 1036
300	5	7-31; 1206	62-77; 1120	98-115; 1219	155-170; 647
		182-206; 1989			
301	1	100-119; 1816			
302	2	109-128; 932	143-162; 2178		
303	4	17-33; 540	54-71; 2700	99-122; 1064	183-203; 2505
304	1	60-72; 1513			
305	3	89-107; 3007	125-143; 1461	174-193; 2228	
306	3	6-34; 1804	48-64; 980	117-132; 599	
307	3	37-52; 1351	67-80; 2411	151-166; 523	
308	11	20-36; 1794	93-108; 1358	118-138; 2196	146-159; 779
		209-223; 2351	294-316; 850	309-325; 967	362-379; 1578
			386-402; 1996	428-454; 1188	462-477; 1965
309	4	25-41; 1707	36-59; 852	61-83; 773	101-120; 1791
310	1	18-35; 2169			
311	4	236-258; 1342	270-285; 1522	304-322; 1138	429-447; 2437
312	1	332-356; 3221			

287

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score			
313	2	17-52; 564	536-556; 3165		
314	2	151-165; 836	427-443; 3134		
315	2	151-165; 836	415-431; 3134		
316	5	56-72; 1759	104-118; 1739	152-181; 3025	199-215; 987
		230-247; 1737			
317	1	438-453; 762			
318	10	44-77; 590	82-97; 1267	160-194; 1095	174-208; 1492
		230-251; 1703	253-278; 1268	287-302; 1352	312-326; 1252
		355-373; 2066	386-403; 1499		
319	4	16-38; 2449	77-94; 1750	109-131; 2443	153-171; 1698
320	7	42-59; 1401	75-99; 1751	110-134; 1209	160-179; 2116
		200-216; 1212	283-296; 2687	319-335; 790	
321	6	16-35; 2306	60-76; 1207	101-115; 1890	155-172; 1646
		201-225; 2512	250-268; 1697		
322	11	89-105; 1259	108-124; 1058	139-157; 1802	168-185; 1278
		189-205; 915	224-240; 1616	311-328; 1587	390-408; 1074
		423-444; 1905	450-468; 1163	552-572; 540	
323	10	11-38; 1993	50-65; 859	106-128; 1632	117-140; 870
		164-184; 1886	194-209; 1335	299-324; 1463	339-352; 930
		413-431; 835	466-481; 1566		
324	1	35-55; 694			
325	1	22-43; 2636			
326	1	152-168; 610			
327	4	22-38; 3134	65-80; 1300	512-531; 2076	542-555; 746
328	3	22-38; 3134	65-80; 1300	493-507; 936	
329	3	22-38; 3134	65-80; 1313	512-531; 2076	
330	4	27-48; 1144	69-92; 2697	119-134; 1835	160-182; 552
331	3	31-47; 1577	652-667; 592	930-952; 3003	
332	1	148-169; 2982			
333	7	83-99; 1049	110-125; 1190	182-198; 1150	206-222; 1406
		232-246; 953	278-295; 1834	338-353; 1407	
334	5	9-35; 1516	26-49; 2339	69-87; 1588	141-155; 2014
		154-180; 579			
335	3	58-73; 589	285-300; 1231	493-509; 2248	
336	8	285-303; 1598	417-430; 866	549-566; 1758	569-583; 995
		634-650; 1821	659-674; 1429	691-709; 2005	724-737; 825
337	1	66-92; 508			
338	7	24-39; 2590	60-73; 600	91-119; 1337	148-163; 566
		196-214; 2187	236-259; 878	272-291; 1508	
339	7	24-39; 2590	60-73; 600	91-119; 1337	148-163; 566
		196-214; 2187	236-259; 878	272-291; 1508	
340	5	18-33; 955	222-237; 670	282-299; 1484	310-325; 786
		710-731; 2486			
341	9	447-464; 826	548-563; 848	646-666; 2709	680-702; 1087
		712-727; 1843	752-770; 1193	799-818; 2230	844-860; 1402
		877-893; 1767			
342	5	25-51; 2632	61-75; 1133	92-120; 1945	141-158; 1186
		177-196; 1468			
343	5	41-59; 1627	54-85; 2078	141-162; 1510	178-199; 2300
		241-266; 1378			
344	7	28-52; 2109	64-85; 1007	95-123; 1859	147-161; 875
		200-219; 1807	247-263; 1555	276-295; 1639	
345	11	91-109; 760	245-262; 900	405-424; 2528	436-454; 1166

288

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score
		460-478; 1710 514-530; 1043 551-573; 2733 597-615; 1300 625-644; 1509 688-707; 1446 773-790; 617
346	10	149-166; 900 309-328; 2528 340-358; 1166 364-382; 1710 418-434; 1043 455-477; 2733 501-519; 1300 529-548; 1509 592-611; 1446 677-694; 617
347	7	38-54; 1710 64-80; 1230 150-169; 1096 177-189; 660 205-220; 1089 247-259; 583 294-311; 1199
348	1	25-44; 1754
349	4	61-78; 1267 92-107; 1758 96-132; 910 125-145; 1211
350	1	63-81; 2993
351	1	21-37; 3067
352	1	33-49; 829
353	1	14-32; 1792
354	1	53-72; 1987
355	1	501-522; 2686
356	2	235-254; 582 307-322; 1905
357	3	305-324; 989 359-385; 512 704-723; 3256
358	1	20-39; 1897
359	1	20-39; 1897
360	1	21-36; 3076
361	2	13-32; 2338 110-126; 621
362	1	342-363; 3126
363	4	25-43; 2055 148-164; 770 232-258; 718 270-283; 1272
364	6	43-59; 1008 80-95; 798 130-149; 886 157-175; 1133 191-212; 1337 226-250; 1425
365	10	58-74; 1806 81-103; 1546 115-127; 710 174-189; 1420 278-299; 1477 321-337; 1182 347-363; 1923 383-398; 1258 403-426; 1703 439-454; 1202
366	3	22-52; 1371 65-89; 1862 100-121; 994
367	1	217-236; 652
368	2	21-36; 2696 95-110; 1111
369	5	576-592; 578 747-762; 2335 764-786; 1265 804-825; 1715 856-871; 1373
370	1	120-140; 3089
371	3	100-115; 939 284-302; 707 332-347; 933
372	7	47-64; 1640 87-101; 700 119-134; 1949 143-159; 507 184-199; 593 208-223; 744 456-477; 2177
373	2	163-175; 1638 182-207; 1865
374	1	32-51; 3413
375	3	225-243; 1004 324-339; 1291 386-402; 1266
376	2	196-214; 1004 313-329; 1173
377	2	126-143; 1381 149-161; 668
378	3	126-143; 1381 149-161; 668 195-220; 807
379	1	80-103; 3414
380	7	20-41; 602 52-71; 1552 83-98; 1700 103-120; 1370 136-151; 2709 162-178; 1788 193-211; 1280
381	3	44-62; 2777 65-80; 1045 141-156; 1507
382	1	92-112; 1518
383	2	73-88; 605 334-356; 1208
384	12	54-69; 1830 90-109; 2293 118-133; 1498 156-176; 884 184-200; 1166 232-251; 1806 282-297; 1680 320-335; 2405 349-364; 1374 377-401; 1798 423-437; 1391 444-463; 2164

289

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score
385	5	49-66; 2934 135-149; 610 177-197; 653 275-289; 698 397-417; 1229
386	5	49-66; 2934 166-188; 504 190-208; 500 266-280; 698 388-408; 1229
387	2	35-61; 782 69-85; 2708
388	2	13-32; 1026 364-383; 1294
389	5	297-315; 565 321-336; 515 340-363; 626 934-954; 875 1131-1147; 556
390	4	27-43; 1142 103-122; 1568 138-154; 868 174-204; 1058
391	3	90-112; 638 127-145; 669 209-229; 733
392	5	195-216; 2012 224-246; 640 258-279; 2594 294-313; 1189 342-362; 2675
393	9	68-88; 2263 115-130; 1131 142-162; 2103 172-187; 986 212-229; 2963 236-251; 1166 274-291; 2044 311- 326; 1229 337-357; 2709
394	1	126-141; 896
395	14	134-159; 1969 296-312; 1030 394-418; 2134 427-440; 1532 432-458; 2248 452-469; 1111 500-518; 1407 536- 549; 1051 616-633; 2001 817-832; 1658 841-858; 2487 866-889; 943 912-934; 1900 940-957; 1433
396	2	311-344; 667 373-390; 788
397	1	204-228; 2681
398	11	61-80; 3083 91-107; 866 120-142; 886 154-169; 1501 196-208; 865 267-286; 1159 315-331; 2009 357- 375; 1205 377-404; 2067 416-433; 913 447-463; 2180
399	2	53-72; 2827 291-307; 809
400	2	28-59; 982 54-69; 843
401	1	188-207; 2756
402	2	120-138; 631 196-211; 534
403	2	64-86; 2717 120-136; 1251
404	6	21-42; 555 76-100; 1949 130-150; 1051 204-219; 943 232-248; 1740 260-278; 1996
405	8	84-101; 750 135-154; 1635 162-178; 1545 187-204; 1038 211-227; 2064 232-245; 1277 265-286; 1440 298- 313; 1011
406	10	167-182; 1236 192-213; 2175 202-237; 869 270-284; 1296 296-316; 1177 309-327; 1613 400-412; 1434 597- 614; 1965 624-660; 681 722-744; 2309
407	1	45-67; 3251
408	3	53-83; 1832 107-121; 1361 128-151; 1826
409	1	165-186; 1496
410	2	328-350; 819 433-448; 634
411	7	26-48; 2329 61-83; 815 95-120; 2154 143-159; 947 205-222; 1700 237-260; 1060 270-292; 1172
412	6	73-87; 1184 104-122; 2026 145-160; 2008 196-215; 2624 235-256; 1873 281-300; 1350
413	2	226-245; 2251 263-287; 800
414	4	48-64; 1636 92-110; 1288 139-157; 930 171-192; 2385
415	10	64-84; 854 188-201; 2590 218-237; 1364 386-401; 2666 405-425; 1179 874-895; 1854 944-961; 1011 1000- 1022; 1158 1040-1065; 894 1072-1088; 1850
416	4	105-120; 2238 127-148; 1679 167-183; 2605 202-217; 1098
417	2	49-64; 631 159-173; 822
418	13	241-255; 643 382-400; 1292 413-428; 1275 433-448; 852

290

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score
		463-485; 1608 491-509; 732 589-605; 1660 630-645; 1543 679-691; 1481 720-735; 2038 775-794; 1386 801-817; 1752 849-864; 1553
419	3	154-172; 1020 185-200; 629 231-251; 1947
420	5	34-50; 668 70-85; 566 264-282; 1020 295-310; 629 341-361; 1947
421	2	18-34; 530 52-73; 703
422	3	208-226; 725 542-558; 567 570-599; 943
423	8	56-71; 578 211-228; 1481 328-346; 644 454-473; 731 587-601; 587 699-714; 553 1039-1055; 612 1489-1518; 771
424	1	411-432; 2031
425	1	51-68; 2943
426	1	106-120; 2492
427	9	42-57; 1250 81-93; 1131 95-111; 1306 103-139; 901 131-148; 1307 160-178; 1366 199-220; 1093 256-276; 1647 311-326; 1736
428	10	42-57; 1250 81-93; 1131 95-111; 1306 103-139; 901 131-148; 1307 160-178; 1366 199-220; 1093 256-276; 1647 314-332; 902 368-384; 990
429	1	85-101; 1852
430	3	198-216; 617 389-404; 1219 429-445; 1499
431	1	42-60; 2634
432	1	215-230; 2143
433	3	29-52; 2263 62-82; 1557 94-113; 2561
434	4	96-112; 1641 167-187; 2265 202-224; 1612 257-272; 2465
435	1	94-114; 2794
436	2	73-92; 2179 123-137; 779
437	1	271-292; 2993
438	1	727-744; 2924
439	1	78-102; 2634
440	4	90-110; 536 114-131; 907 183-195; 654 268-291; 977
441	4	90-110; 536 114-131; 907 183-195; 654 268-291; 977
442	4	90-110; 536 114-131; 907 183-195; 654 268-291; 977
443	5	53-69; 2297 83-98; 1058 145-163; 1504 179-194; 1353 206-222; 2021
444	3	78-98; 2028 134-150; 1060 224-243; 1701
445	4	17-42; 706 53-70; 1592 97-112; 1041 142-160; 2123
446	4	198-214; 755 274-289; 868 306-321; 1260 330-345; 737
447	1	46-64; 1815
448	1	129-154; 569
449	1	468-489; 2129
450	1	354-373; 3038
451	2	64-79; 726 73-97; 888
452	3	151-166; 645 186-208; 1300 255-270; 508
453	3	82-95; 530 112-129; 1374 1470-1491; 3847
454	2	30-43; 2002 302-320; 1525
455	2	84-96; 576 892-911; 2528
456	1	28-48; 1700
457	1	77-103; 2678
458	5	25-50; 2582 61-82; 1050 92-120; 827 140-155; 831 199-214; 1366
459	7	33-50; 2479 58-73; 1393 94-115; 882 144-162; 671

291

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score
		214-231; 2323 295-309; 1593 379-398; 2767
460	2	39-58; 1574 90-107; 2845
461	2	166-183; 1505 206-228; 2412
462	2	103-118; 554 158-176; 1691
463	4	155-170; 1480 316-331; 707 340-357; 1159 368-381; 609
464	2	63-79; 1054 638-658; 2381
465	1	94-109; 1151
466	3	340-355; 673 386-400; 599 435-451; 1027
467	2	40-55; 884 74-88; 904
468	3	63-87; 668 134-150; 782 165-182; 1034
469	10	49-66; 1360 79-94; 1389 111-124; 917 138-153; 1267 165-179; 890 182-202; 532 229-243; 898 254- 271; 1978 270-288; 1076 309-325; 1735
470	3	107-122; 720 141-162; 1315 193-208; 759
471	2	146-161; 510 194-221; 1018
472	3	16-32; 1307 69-83; 1789 88-114; 1279
473	4	16-32; 1307 69-83; 1789 88-114; 1279 129-154; 1198
474	4	38-54; 1155 103-121; 2670 134-148; 1558 195-215; 1883
475	5	90-112; 638 127-145; 669 209-229; 749 313-331; 644 406-422; 904
476	2	337-361; 1379 527-543; 559
477	6	28-43; 1439 94-123; 768 143-157; 1354 200-222; 2716 240-263; 1191 273-295; 1338
478	4	71-88; 2706 116-137; 867 136-153; 1128 171-195; 863
479	4	47-59; 1552 63-86; 2366 107-124; 1545 143-170; 2265
480	4	27-60; 710 83-101; 931 116-152; 668 603-627; 1141
481	13	265-279; 643 417-435; 1292 448-463; 1319 468-483; 852 498-520; 1608 526-544; 732 627-643; 1660 668- 683; 1543 717-729; 1481 758-773; 2038 813-832; 1386 839-855; 1752 887-902; 1553
482	5	37-50; 569 445-463; 2049 489-513; 1074 529-549; 2945 552-570; 1394
483	5	37-53; 1814 71-86; 1511 93-108; 1516 121-136; 1562 160-175; 2012
484	1	103-118; 1952
485	6	121-139; 864 584-605; 2969 619-635; 1436 649-667; 1359 699-719; 1257 746-762; 1819
486	7	17-40; 2341 55-70; 1212 90-111; 1353 132-152; 1570 185-203; 1862 221-237; 1592 258-281; 755
487	1	73-92; 1951
488	2	65-80; 2366 89-102; 1530
490	3	62-76; 1511 91-109; 609 160-185; 629
491	7	25-40; 1285 58-76; 922 91-107; 584 142-164; 1715 200-218; 1486 244-259; 2257 272-284; 1020
492	2	159-174; 702 216-234; 2518
493	3	20-35; 506 49-69; 984 333-352; 1717
494	1	363-379; 1359
495	9	52-71; 2689 88-103; 1366 153-165; 2603 188-205; 1124 221-240; 2123 267-279; 1245 290-309; 1070 323- 337; 1257 345-359; 844
496	2	151-166; 1709 214-235; 1665
497	6	102-119; 577 136-153; 1288 149-173; 551 194-212; 697 262-281; 1364 304-316; 1698



292

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score
498	2	136-151; 751 193-212; 2670
499	7	181-196; 658 272-287; 862 740-753; 1177 827-845; 521 900-920; 771 926-941; 1124 1467-1492; 835
500	2	26-42; 553 172-188; 2514
501	1	451-466; 826
502	6	24-45; 1693 72-84; 881 95-114; 996 141-153; 878 200-220; 2700 251-265; 1354
503	6	726-747; 724 776-791; 985 806-828; 806 1019-1039; 680 1058-1082; 605 1111-1131; 929
504	2	73-89; 1003 572-595; 2977
505	7	68-91; 2217 103-117; 1024 145-162; 1476 184-200; 1937 239-258; 2428 287-302; 1125 312-334; 1293
506	4	59-74; 784 411-426; 543 555-570; 1432 755-770; 543
507	5	48-71; 2145 138-154; 508 233-257; 580 278-290; 793 341-362; 1028
508	4	22-41; 661 753-771; 682 866-881; 639 948-965; 1707
509	2	93-109; 2922 246-262; 610
510	3	45-71; 1224 97-119; 2200 105-128; 1270
511	1	96-118; 2253
512	1	213-228; 2903
513	12	27-53; 2787 63-76; 997 108-129; 707 155-170; 1049 201-221; 1704 247-263; 1270 274-296; 1442 385- 397; 1137 437-452; 1414 510-529; 799 549-563; 1638 576-596; 953
514	8	200-215; 1460 271-289; 2381 361-378; 1369 396-416; 2113 440-455; 1279 477-495; 1320 521-541; 1573 573- 593; 2337
515	6	94-111; 2450 116-137; 985 152-171; 2459 188-203; 1343 223-243; 1668 254-269; 1184
516	7	422-439; 2505 460-482; 954 494-527; 1524 546-562; 1289 588-606; 2147 631-648; 1264 667-686; 1796
517	2	23-36; 582 40-73; 1069
518	11	20-35; 1776 53-68; 1782 86-102; 1155 131-146; 1074 164-179; 2382 442-459; 1328 495-510; 1765 527- 542; 1214 547-562; 1720 590-617; 795 625-644; 1995
519	9	314-331; 826 415-430; 848 513-533; 2709 547-569; 1087 579-594; 1843 619-637; 1193 666-685; 2230 711- 727; 1402 744-760; 1767
520	2	62-77; 645 116-133; 1910
521	5	70-85; 975 101-119; 2374 140-158; 1457 228-244; 2107 256-274; 1074
522	7	81-97; 2470 121-136; 1224 149-176; 1604 209-225; 1439 267-286; 2119 309-324; 1473 376-393; 1898
523	2	34-48; 680 160-175; 848
524	7	59-83; 2997 95-116; 1032 141-156; 1091 175-192; 1755 228-249; 1807 281-297; 1698 318-341; 1040
525	3	34-52; 2348 155-170; 575 323-337; 2673
526	5	65-83; 3178 93-107; 1020 137-158; 2389 172-192; 1494 224-241; 3165
527	7	38-55; 2045 125-140; 1136 320-339; 2947 335-360; 1228 364-386; 1097 422-437; 943 451-469; 1867
528	11	118-133; 2943 199-212; 1121 230-251; 2184 264-285; 1606 302-317; 1270 343-360; 1239 422-446; 1581 457- 472; 1460 492-511; 2540 503-532; 504 562-577; 1749

293

Table 8

SEQ ID NO:	Number of Transmembrane Domains	For Each Transmembrane Domain, its Transmembrane Domain Position in SEQ ID NO: and its TM Pred Score
529	4	81-108; 674 150-166; 1423 300-315; 1978 486-501; 799
530	6	27-43; 974 66-85; 1887 98-114; 1177 120-142; 1864 163-180; 871 208-225; 2625
531	4	88-104; 2727 112-137; 1466 152-173; 1863 195-216; 1523
532	8	55-71; 2368 82-96; 847 117-141; 1703 161-180; 1265 218-237; 2278 265-281; 1248 297-313; 748 325-346; 1097
533	3	471-484; 505 578-593; 1235 605-619; 981
534	10	50-67; 900 188-207; 2528 219-237; 1166 243-261; 1710 297-313; 1043 334-356; 2733 380-398; 1300 408-427; 1509 471-490; 1446 556-573; 617
535	7	410-425; 2180 656-671; 1017 692-711; 1695 717-735; 898 751-767; 2256 773-789; 1341 809-824; 2908
536	7	433-448; 2180 679-694; 1017 715-734; 1695 740-758; 898 774-790; 2256 796-812; 1341 832-847; 2908
537	1	66-88; 2934
538	7	26-51; 1782 61-83; 603 91-120; 1188 140-154; 1223 198-226; 2284 245-260; 1580 273-292; 1207
539	7	27-39; 1172 50-65; 1681 80-104; 1084 109-138; 1616 151-163; 1311 165-188; 1247 200-215; 971
540	3	29-52; 2263 62-82; 1557 94-113; 2561
541	2	100-116; 1881 135-156; 1002
542	3	126-145; 939 142-165; 508 680-701; 2775
543	1	26-44; 863
544	1	83-99; 2738
545	11	25-40; 737 250-267; 2877 277-299; 1267 325-342; 1801 357-370; 1156 440-459; 2243 702-720; 1515 729-746; 2454 755-770; 589 799-821; 2411 836-850; 1194
546	6	30-46; 1302 49-69; 1510 76-90; 1070 104-123; 1711 147-160; 1419 186-202; 2239
547	5	55-70; 1001 95-117; 1013 386-406; 973 664-682; 599 1655-1668; 1126
548	1	82-101; 3223
549	3	55-73; 2750 79-96; 1280 115-129; 1733
550	8	25-48; 2164 61-75; 774 91-120; 1887 140-158; 937 199-219; 2862 245-260; 1258 273-292; 1715 330-345; 782
551	13	334-354; 586 480-495; 1208 509-529; 1145 565-581; 1273 593-611; 1007 695-710; 1443 730-748; 1753 784-800; 1657 826-846; 2236 882-900; 1281 885-913; 1566 902-926; 923 972-989; 1888
552	9	54-76; 2605 103-118; 984 130-150; 2154 160-175; 1065 199-216; 3177 225-239; 1416 262-282; 1291 299-314; 1383 325-342; 2377

294

Table 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
1	277	553	773	790 11261
2	278	554	774	790 11261
3	279	555	775	790 11261
4	280	556	776	784 4082
5	281	557	777	784 7871
6	282			
7	283			
8	284	558	778	785 2318
9	285	559	779	784 5413
10	286	560	780	785 3232
11	287	561	781	790 89
12	288	562	782	787 5259
13	289	563	783	785 1914
14	290			
15	291	564	784	785 1259
16	292			
17	293			
18	294	565	785	789 3965
19	295	566	786	785 3694
20	296	567	787	787 4872
21	297	568	788	787 9713
22	298	569	789	787 2349
23	299	570	790	785 1465
24	300	571	791	784 3151
25	301	572	792	787 8974
26	302	573	793	790 7111
27	303	574	794	787 2905
28	304	575	795	784 7871
29	305	576	796	791 2843
30	306	577	797	784 9890
31	307	578	798	790 10356
32	308	579	799	784 2633
33	309	580	800	790 3779
34	310			
35	311	581	801	784 2684
36	312	582	802	784 5473
37	313	583	803	785 332
38	314	584	804	784 8092
39	315	585	805	784 8092
40	316	586	806	787 2275
41	317	587	807	784 4451
42	318	588	808	784 8006
43	319	589	809	785 769
44	320			
45	321	590	810	787 4983
46	322	591	811	787 9291
47	323	592	812	785 1000
48	324			
49	325			

295

Table 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
50	326			
51	327	593	813	787 3917
52	328	594	814	787 3917
53	329	595	815	787 3917
54	330	596	816	790 14759
55	331	597	817	784 1652
56	332	598	818	787 10209
57	333	599	819	784 3955
58	334	600	820	784 7153
59	335			
60	336	601	821	784 3946
61	337	602	822	789 3723
62	338	603	823	787 3770
63	339	604	824	787 3770
64	340	605	825	784 2336
65	341	606	826	789 4217
66	342			
67	343			
68	344			
69	345	607	827	785 1541
70	346	608	828	785 1541
71	347			
72	348	609	829	784 3641
73	349			
74	350	610	830	785 2572
75	351			
76	352	611	831	784 6671
77	353			
78	354	612	832	784 7805
79	355	613	833	785 2923
80	356	614	834	784 5115
81	357	615	835	784 1141
82	358	616	836	784 2449
83	359	617	837	784 2449
84	360	618	838	788 13754
85	361			
86	362	619	839	784 8759
87	363	620	840	785 842
88	364	621	841	784 1145
89	365	622	842	784 10001
90	366	623	843	784 6967
91	367	624	844	787 5991
92	368	625	845	787 3955
93	369	626	846	784 5413
94	370	627	847	785 749
95	371	628	848	784 7384
96	372	629	849	784 3517
97	373	630	850	784 9490
98	374	631	851	785 442
99	375	632	852	791 16

296

Table 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
100	376	633	853	791 16
101	377	634	854	790 26559
102	378	635	855	790 26559
103	379	636	856	787 9546
104	380	637	857	784 6047
105	381	638	858	784 2820
106	382	639	859	784 3402
107	383	640	860	784 5142
108	384	641	861	784 4630
109	385	642	862	787 1021
110	386	643	863	787 1021
111	387	644	864	784 4543
112	388	645	865	787 4613
113	389	646	866	784 1107
114	390	647	867	790 14636
115	391	648	868	787 3544
116	392	649	869	784 2281
117	393	650	870	784 4265
118	394			
119	395	651	871	784 1885
120	396	652	872	790 2819
121	397	653	873	784 7981
122	398	654	874	785 2923
123	399	655	875	784 4589
124	400			
125	401	656	876	790 26407
126	402	657	877	790 8012
127	403	658	878	791 131
128	404	659	879	790 16319
129	405	660	880	790 18649
130	406	661	881	789 4901
131	407			
132	408	662	882	784 4813
133	409			
134	410	663	883	784 3977
135	411	664	884	784 3507
136	412	665	885	784 8101
137	413	666	886	784 1263
138	414	667	887	791 3081
139	415	668	888	792 5307
140	416	669	889	784 337
141	417	670	890	790 311
142	418	671	891	784 3298
143	419	672	892	788 2631
144	420	673	893	788 2631
145	421			
146	422	674	894	787 2204
147	423	675	895	787 4220
148	424	676	896	784 1948
149	425	677	897	791 2929

297

Table 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
150	426	678	898	785 86
151	427	679	899	784 4387
152	428	680	900	784 4387
153	429			
154	430	681	901	790 26525
155	431			
156	432			
157	433	682	902	784 6050
158	434			
159	435	683	903	784 5883
160	436			
161	437	684	904	784 1866
162	438	685	905	784 623
163	439	686	906	784 2034
164	440	687	907	784 2132
165	441	688	908	784 2132
166	442	689	909	784 2132
167	443	690	910	787 2259
168	444	691	911	784 5922
169	445	692	912	784 5356
170	446			
171	447	693	913	784 2543
172	448	694	914	784 4218
173	449	695	915	784 2452
174	450	696	916	784 3125
175	451			
176	452			
177	453	697	917	787 5429
178	454	698	918	789 3376
179	455			
180	456	699	919	787 7913
181	457	700	920	790 26693
182	458	701	921	787 4277
183	459			
184	460	702	922	784 722
185	461			
186	462	703	923	787 5679
187	463	704	924	784 1990
188	464	705	925	784 3590
189	465	706	926	787 242
190	466	707	927	784 10036
191	467			
192	468	708	928	784 3120
193	469			
194	470	709	929	784 4715
195	471	710	930	790 10323
196	472	711	931	784 8845
197	473			
198	474			
199	475	712	932	790 13184

298

Table 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
200	476	713	933	787 9837
201	477	714	934	790 27173
202	478	715	935	787 5608
203	479	716	936	784 1000
204	480			
205	481	717	937	784 3298
206	482	718	938	787 2264
207	483	719	939	787 9869
208	484			
209	485	720	940	784 8003
210	486	721	941	784 4891
211	487	722	942	784 220
212	488	723	943	784 3720
213	489	724	944	784 8022
214	490	725	945	784 3117
215	491			
216	492	726	946	792 6338
217	493	727	947	790 16986
218	494			
219	495	728	948	785 3255
220	496			
221	497	729	949	784 2248
222	498	730	950	790 25345
223	499	731	951	784 5062
224	500	732	952	789 817
225	501			
226	502	733	953	787 8810
227	503	734	954	787 1572
228	504	735	955	790 12296
229	505	736	956	790 27173
230	506	737	957	784 1571
231	507	738	958	784 3746
232	508	739	959	784 1097
233	509			
234	510			
235	511	740	960	784 5926
236	512			
237	513			
238	514	741	961	784 5318
239	515	742	962	790 12758
240	516	743	963	784 5328
241	517			
242	518	744	964	785 507
243	519	745	965	789 4217
244	520	746	966	791 2641
245	521	747	967	790 23507
246	522	748	968	784 2608
247	523	749	969	787 84
248	524	750	970	790 16983
249	525			

299

Table 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
250	526			
251	527			
252	528	751	971	787 4538
253	529	752	972	784 4452
254	530	753	973	784 3405
255	531	754	974	787 2752
256	532			
257	533			
258	534	755	975	785 1541
259	535	756	976	784 4406
260	536	757	977	784 4406
261	537	758	978	785 33
262	538	759	979	787 5204
263	539	760	980	784 482
264	540	761	981	787 6564
265	541	762	982	788 6847
266	542	763	983	785 1239
267	543	764	984	784 4069
268	544	765	985	785 1321
269	545	766	986	785 658
270	546	767	987	787 3324
271	547	768	988	784 10120
272	548	769	989	787 10039
273	549	770	990	787 9881
274	550			
275	551	771	991	789 1858
276	552	772	992	784 10115

\*784\_XXX = SEQ ID NO: XXX of Attorney Docket No. 784, US Serial No. 09/488,725 filed 01/21/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

785\_XXX = SEQ ID NO: XXX of Attorney Docket No. 785, US Serial No. 09/491,404 filed 01/25/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

787\_XXX = SEQ ID NO: XXX of Attorney Docket No. 787, US Serial No. 09/496,914 filed 02/03/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

788\_XXX = SEQ ID NO: XXX of Attorney Docket No. 788, US Serial No. 09/515,126 filed 02/28/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

789\_XXX = SEQ ID NO: XXX of Attorney Docket No. 789, US Serial No. 09/519,705 filed 03/07/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.



300  
Table 9

790\_XXX = SEQ ID NO: XXX of Attorney Docket No. 790, US Serial No. 09/540,217 filed 03/31/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

791\_XXX = SEQ ID NO: XXX of Attorney Docket No. 791, US Serial No. 09/552,929 filed 04/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

792\_XXX = SEQ ID NO: XXX of Attorney Docket No. 792, US Serial No. 09/577,408 filed 05/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

301  
Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/323,739
1	277	1
2	278	2
3	279	3
4	280	4
5	281	5
6	282	6
7	283	7
8	284	8
9	285	9
10	286	10
11	287	11
12	288	12
13	289	13
14	290	14
15	291	15
16	292	16
17	293	17
18	294	18
19	295	19
20	296	20
21	297	21
22	298	22
23	299	23
24	300	24
25	301	25
26	302	26
27	303	27
28	304	28
29	305	29
30	306	30
31	307	31
32	308	32
33	309	33
34	310	34
35	311	35
36	312	36
37	313	37
38	314	38
39	315	39
40	316	40
41	317	41
42	318	42
43	319	43
44	320	44
45	321	45
46	322	46
47	323	47
48	324	48
49	325	49
50	326	50
51	327	51
52	328	52

302

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/323,739
53	329	53
54	330	54
55	331	55
56	332	56
57	333	57
58	334	58
59	335	59
60	336	60
61	337	61
62	338	62
63	339	63
64	340	64
65	341	65
66	342	66
67	343	67
68	344	68
69	345	69
70	346	70
71	347	71
72	348	72
73	349	73
74	350	74
75	351	75
76	352	76
77	353	77
78	354	78
79	355	79
80	356	80
81	357	81
82	358	82
83	359	83
84	360	84
85	361	85
86	362	86
87	363	87
88	364	88
89	365	89
90	366	90
91	367	91
92	368	92
93	369	93
94	370	94
95	371	95
96	372	96
97	373	97
98	374	98
99	375	99
100	376	100
101	377	101
102	378	102
103	379	103
104	380	104
105	381	105

303

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/323,739
106	382	106
107	383	107
108	384	108
109	385	109
110	386	110
111	387	111
112	388	112
113	389	113
114	390	114
115	391	115
116	392	116
117	393	117
118	394	118
119	395	119
120	396	120
121	397	121
122	398	122
123	399	123
124	400	124
125	401	125
126	402	126
127	403	127
128	404	128
129	405	129
130	406	130
131	407	131
132	408	132
133	409	133
134	410	134
135	411	135
136	412	136
137	413	137
138	414	138
139	415	139
140	416	140
141	417	141
142	418	142
143	419	143
144	420	144
145	421	145
146	422	146
147	423	147
148	424	148
149	425	149
150	426	150
151	427	151
152	428	152
153	429	153
154	430	154
155	431	155
156	432	156
157	433	157
158	434	158

304

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/323,739
159	435	159
160	436	160
161	437	161
162	438	162
163	439	163
164	440	164
165	441	165
166	442	166
167	443	167
168	444	168
169	445	169
170	446	170
171	447	171
172	448	172
173	449	173
174	450	174
175	451	175
176	452	176
177	453	177
178	454	178
179	455	179
180	456	180
181	457	181
182	458	182
183	459	183
184	460	184
185	461	185
186	462	186
187	463	187
188	464	188
189	465	189
190	466	190
191	467	191
192	468	192
193	469	193
194	470	194
195	471	195
196	472	196
197	473	197
198	474	198
199	475	199
200	476	200
201	477	201
202	478	202
203	479	203
204	480	204
205	481	205
206	482	206
207	483	207
208	484	208
209	485	209
210	486	210
211	487	211

305

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/323,739
212	488	212
213	489	213
214	490	214
215	491	215
216	492	216
217	493	217
218	494	218
219	495	219
220	496	220
221	497	221
222	498	222
223	499	223
224	500	224
225	501	225
226	502	226
227	503	227
228	504	228
229	505	229
230	506	230
231	507	231
232	508	232
233	509	233
234	510	234
235	511	235
236	512	236
237	513	237
238	514	238
239	515	239
240	516	240
241	517	241
242	518	242
243	519	243
244	520	244
245	521	245
246	522	246
247	523	247
248	524	248
249	525	249
250	526	250
251	527	251
252	528	252
253	529	253
254	530	254
255	531	255
256	532	256
257	533	257
258	534	258
259	535	259
260	536	260
261	537	261
262	538	262
263	539	263
264	540	264

306

Table 10

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/323,739
265	541	265
266	542	266
267	543	267
268	544	268
269	545	269
270	546	270
271	547	271
272	548	272
273	549	273
274	550	274
275	551	275
276	552	276